On Tackling Fundamental Constraints in Brain-Computer Interface Decoding via Deep Neural Networks

NIK-AZNAN, NIK,KHADIJAH,BINTI

On Tackling Fundamental Constraints in Brain-Computer Interface Decoding via Deep Neural Networks

Nik Khadijah Nik Aznan

A thesis presented for the degree of
Doctor of Philosophy at Durham University

Department of Computer Science
Durham University
United Kingdom

15th June 2021

On Tackling Fundamental Constraints in Brain-Computer Interface Decoding via Deep Neural Networks

Nik Khadijah Nik Aznan

Submitted for the degree of Doctor of Philosophy
15th June 2021

A Brain-Computer Interface (BCI) is a system that provides a communication and control medium between human cortical signals and external devices, with the primary aim to assist or to be used by patients who suffer from a neuromuscular disease. Despite significant recent progress in the area of BCI, there are numerous shortcomings associated with decoding Electroencephalography-based BCI signals in real-world environments. These include, but are not limited to, the cumbersome nature of the equipment, complications in collecting large quantities of real-world data, the rigid experimentation protocol and the challenges of accurate signal decoding, especially in making a system work in real-time. Hence, the core purpose of this work is to investigate improving the applicability and usability of BCI systems, whilst preserving signal decoding accuracy.

Recent advances in Deep Neural Networks (DNN) provide the possibility for signal processing to automatically learn the best representation of a signal, contributing to improved performance even with a noisy input signal. Subsequently, this thesis focuses on the use of novel DNN-based approaches for tackling some of the key underlying constraints within the area of BCI. For example, recent technological improvements in acquisition hardware have made it possible to eliminate the pre-existing rigid experimentation procedure, albeit resulting in noisier signal capture. However, through the use of a DNN-based model, it is possible to preserve the accuracy of the predictions from the decoded signals. Moreover, this research demonstrates that by leveraging DNN-based image and signal understanding, it is feasible to facilitate real-time BCI applications in a natural environment. Additionally, the capability of DNN to generate realistic synthetic data is shown to be a potential solution in reducing the requirement for costly data collection. Work is also performed in addressing the well-known issues regarding subject bias in BCI models by generating data with reduced subject-specific features.

The overall contribution of this thesis is to address the key fundamental limitations of BCI systems. This includes the unyielding traditional experimentation procedure, the mandatory extended calibration stage and sustaining accurate signal decoding in real-time. These limitations lead to a fragile BCI system that is demanding to use and only suited for deployment in a controlled laboratory. Overall contributions of this research aim to improve the robustness of BCI systems and enable new applications for use in the real-world.

Declaration

The work in this thesis is based on research carried out within the Department of Computer Science, Department of Engineering and Department of Psychology, Durham University, UK. No part of this thesis has been submitted elsewhere for any other degree or qualification, and it is all the author's work unless referenced to the contrary in the text.

Acknowledgements

I owe my deepest gratitude to my supervisor, Prof Toby Breckon for not only giving me this opportunity to work on very interesting research, but for giving his precious time and effort towards my research – constantly pouring in amazing ideas and constructive comments.

I would like to express my appreciation to Dr Jason Connolly for his enthusiasm and willingness in assisting me especially as a Neuroscientist.

During the course of this degree, I was lucky enough to meet my partner, Dr Stephen Bonner – thank you for your invaluable support and for always being there for me.

I would like to express my gratitude to my parents Mrs Norizan Kasim and Mr Nik Aznan Nik Abdullah for their constant support, endless prayers, and encouragement since day one of me applying to further my study. Special thanks to my UK family Dr Louise Bonner and Mr Neil Bonner for always taking care of me and offering invaluable help.

Sincere thanks to all my friends; Dr Nur Afiqah Sukiran, Dr Nor Mazlin Zahari, Dr Amir Atapour-Abarghouei, Dr Philip Jackson and Josh Podmore who have provided invaluable help and support throughout the years.

A big thanks to all the participants who took part in my research experiments.

Finally, I would like to thank the Ministry of Higher Education Malaysia and Universiti Teknikal Malaysia Melaka for giving me this opportunity and sponsoring this research.

"We tell our story,
not for the purposes of pleasure,
but for the advancement of knowledge,
and for the good of us all."
- James Glaisher, The Aeronouts 2019

Contents

Abstract i

Declaration iii

Acknowledgements iv

Contents vi

List of Figures x

List of Tables xiv

List of Algorithms xvi

List of Abbreviations xvii

1 Introduction 1

1.1 Research Questions and Motivation . . . 2
1.2 Research Aim and Objectives . . . 5
1.3 Thesis Contribution . . . 6
1.4 Publications . . . 7
1.5 Scope . . . 8
1.6 Thesis Structure . . . 9

2 Literature Review 11

2.1 BCI Signal Acquisition . . . 11
2.1.1 EEG-based BCI . . . 12
2.1.2 Dry-EEG vs Wet-EEG System . . . 14

2.1.3 Cognionics Dry-EEG System . . . 16
2.2 BCI Neuro-physiological Responses . . . 18
2.2.1 Type of Responses . . . 18
2.2.2 SSVEP-based BCI . . . 20
2.3 Bio-Signals Pattern Recognition . . . 27
2.3.1 Traditional Machine Learning Approaches . . . 28
2.3.2 Deep Neural Network Approach . . . 30
2.3.3 Bio-Signals Pattern Generation . . . 35
2.4 Limitations of BCI . . . 37
2.5 Relevance to Contributions . . . 38

3 Classifying Dry-EEG Bio-signals **41**
3.1 Introduction . . . 42
3.2 Related Work . . . 43
3.3 Experimental Setup . . . 44
3.3.1 Dry-EEG Headset . . . 45
3.3.2 SSVEP Stimuli . . . 47
3.3.3 Data Collection . . . 49
3.4 Model Overview and Evaluation Methodology . . . 53
3.4.1 SSVEP Convolutional Unit . . . 53
3.4.2 Traditional Machine Learning . . . 56
3.4.3 Recurrent Neural Network . . . 58
3.5 Results and Discussion . . . 59
3.5.1 SSVEP Classification on Subject S01 . . . 60
3.5.2 SSVEP Classification on Multiple Subject . . . 61
3.5.3 SSVEP Classification Across Subjects . . . 62
3.5.4 Generalisation Capability to an Unseen Subject . . . 62
3.6 Conclusion . . . 64

4 SSVEP-based Variable Natural Environment Stimuli **67**
4.1 Introduction . . . 68
4.2 Related Work . . . 70
4.3 Variable BCI Stimuli . . . 72
4.3.1 Single Shot MultiBox Object Detector . . . 72
4.3.2 Stimuli Setup . . . 73

4.4 Methodology . . . 73
4.4.1 Calibration Stage . . . 75
4.4.2 SSVEP Classification . . . 76
4.4.3 Robot Navigation . . . 77
4.4.4 Offline Dataset . . . 78
4.5 Results and Discussion . . . 81
4.5.1 Calibration Statistical Performance . . . 81
4.5.2 On-line Real-time Performance . . . 82
4.5.3 Offline Dataset Performance . . . 86
4.6 Conclusion . . . 90

5 Generating Brain Signals **92**
5.1 Introduction . . . 93
5.2 Related Work . . . 95
5.2.1 Neural-Based Generative Models . . . 96
5.2.2 Literature Review . . . 98
5.3 Proposed Generative Networks . . . 100
5.3.1 Deep Convolutional Generative Adversarial Network . . . 100
5.3.2 Improved Wasserstein Generative Adversarial Network . . . 102
5.3.3 Variational Autoencoder . . . 103
5.4 Experimental Setup . . . 104
5.4.1 Empirical SSVEP Dry-EEG Datasets . . . 104
5.4.2 Synthetic Data Generation Procedure . . . 106
5.4.3 Classification Procedure . . . 106
5.5 Results . . . 107
5.5.1 Data Visualization via Fast Fourier Transforms . . . 107
5.5.2 Mixed Real and Synthetic Data Classification . . . 109
5.5.3 Classification with Pre-Training . . . 110
5.5.4 Cross-Subject Generalisation . . . 112
5.6 Conclusion . . . 113

6 Synthetic EEG Signals via Subject Invariant GAN **119**
6.1 Introduction . . . 120
6.2 Related Work . . . 123
6.3 Proposed Approach . . . 126

6.3.1 Generator and Discriminator Networks . . . 126
6.3.2 Auxiliary Classifier Network . . . 127
6.3.3 Creating Subject Invariant EEG Signals . . . 129
6.3.4 Implementation Details . . . 130
6.4 Experimental Setup . . . 130
6.4.1 Datasets . . . 130
6.4.2 Evaluation Methodology . . . 132
6.5 Experimental Results . . . 135
6.5.1 SSVEP Classification for a Single Subject . . . 136
6.5.2 Subject-biometric Classification . . . 137
6.5.3 SSVEP Classification for Unseen Subject : Leave-One-Out . . . 138
6.5.4 SSVEP Classification for Unseen Subject : Cross-Task . . . 141
6.6 Conclusion . . . 145

7 Conclusion **147**
7.1 Review of Key Contributions . . . 148
7.2 Review of Research Aim and Objectives . . . 151
7.3 Limitations . . . 151
7.4 Future Work . . . 152
7.4.1 Improvements to Current Work . . . 153
7.4.2 Expansions to the Work . . . 153

References **156**

A Data Acquisition **183**

B Variable Stimuli **187**

List of Figures

1.1 Block diagram of the main BCI components 1

2.1 Example of wet EEG technology and dry EEG technology 13
2.2 Traditional BCI signal processing pipeline 28
2.3 Example of Deep Neural Network pipeline on BCI signal processing via Convolutional Neural Network 32

3.1 Cognionics Inc Quick-20 dry-EEG sensor channels (highlighted in blue) 46
3.2 Two different patterns are drawn on a PsychoPy window 47
3.3 Integration set-up between dry-EEG headset and SSVEP stimuli for bio-signals data collection. 50
3.4 Illustrative raw signal data as captured from dry-EEG 51
3.5 FFT plots on four frequency class show the peak frequencies and the harmonics of every classes respectively 52
3.6 1D Convolutional architecture for dry-EEG signals 54
3.7 Our proposed 1D CNN architecture including our proposed SSVEP Convolutional Unit (SCU, highlighted in pink) 55
3.8 RNN sequences used to classify dry-EEG bio-signals 58
3.9 SCU classifying on single subject (S01) for frequency classes (Hz), highlighting the per-class statistical accuracy (Maximal result being accuracy = 1.0 in the matrix diagonals) 61
3.10 Confusion matrices for the SSVEP classification across subjects for frequency classes (Hz), highlighting the per-class statistical accuracy (Maximal result being accuracy = 1.0 in the matrix diagonals). 63
3.11 Our deeper CNN architecture for *unseen* subjects 64

3.12 Confusion matrices for the SSVEP classification on unseen subject (S04) for frequency classes (Hz) by using Deep CNN highlighting the per-class statistical accuracy (Maximal result being accuracy = 1.0 in the matrix diagonals) 65

4.1 Variable SSVEP stimuli based on object detection flickering (left column) to navigate the robot towards the object and navigational arrow flickering (right column) to move the robot facing a new environmental scene. 74
4.2 Overview of the experimental approach proposed. 75
4.3 The 1D CNN architecture used to classify the EEG signals for both offline dataset and real-time experiments (SCU, highlighted in pink). 76
4.4 Flowchart of real-time robot navigation. 79
4.5 Navigation plan for real-time experimentation. 80
4.6 Confusion matrices for the classification of real-time EEG signals during the robot navigation (maximal result being *accuracy* = 1.0 in the matrix diagonals). 83
4.7 Sample of humanoid robot navigation during a real-time experiment. 85
4.8 FFT when taking across samples on all nine subjects. The plots show the peak frequency and the harmonics of all three frequencies. 88

5.1 Details of the collection process for an SSVEP EEG dataset (*Video-Stimuli*) using a video based stimuli. 94
5.2 An overview of the generative adversarial network. 100
5.3 Our proposed GAN. 101
5.4 An overview of the Variational Auto-encoder. 103
5.5 Flowchart detailing the recording procedure for the *Online* and *Offline* SSVEP EEG Datasets. The highlighted region of the figure shows the humanoid robot used to create the *Online dataset*. 105
5.6 The SCU architecture used to classify the EEG signals. 107
5.7 Comparing real and synthetic data from the generative models. Synthetic data clearly displays the characteristic SSVEP frequency peaks at the same frequencies as those observed in the real data. 109
5.8 Flowchart for Pre-training classification experiment. 111
5.9 Convergence of the Cross-Entropy value plotted over training epochs for models with and without pre-training on synthetic data. 113
5.10 Test accuracy when varying the volume of synthetic data used for pre-training. Doted line indicates 500 samples. 115

5.11 Pre-training time 116

5.12 Flowchart for Training on data from different subjects. 117

6.1 Teleoperation task with the required calibration stage identified in the red dotted block. 121

6.2 Confusion matrix for subject classification accuracy (Maximal result being accuracy = 1.0 on the diagonal). 122

6.3 The Deep Convolutional Generative Adversarial Network (DC-GAN) architecture, where the Generator (G) is trained to fool the Discriminator (D). 127

6.4 The Auxiliary Classifier Generative Adversarial Network (AC-GAN) architecture, where the Generator (G) is trained to produce data which can fool the Discriminator (D) and to produce data which can be classified as belonging to a certain class by the Auxiliary network (Aux). 128

6.5 Our Subject Invariant SSVEP Generative Adversarial Network. The Generator (G) produces data with subject-specific information removed (Sub) that can fool the Discriminator (D) and is classified as a certain frequency (Aux). 131

6.6 Evaluation method for SSVEP classification for single subject, where the training and testing data are from the same subject. Dotted lines indicate SSVEP classification. 133

6.7 Evaluation method for SSVEP classification for leave-one-out unseen subject where the test data is from an unseen subject S0x with no prior training. Frozen subject-biometric classifier is used to train SIS-GAN. Dotted lines indicate SSVEP classification. 135

6.8 Evaluation method for SSVEP classification for cross-task unseen subject where the test data is from the online dataset with no prior training. Frozen subject-biometric classifier is used to train SIS-GAN. Dotted lines indicate SSVEP classification. 136

6.9 The distribution of data points embedding using T-distributed Stochastic Neighbor Embedding. 139

6.10 Comparing Fast-Fourier Transforms (FFT) of real and synthetic data from all the generative models. The Synthetic data clearly displays the characteristic SSVEP frequency peaks and associated harmonics. 143

6.11 Softmax probability values taken from the pre-trained subject-biometric classification network for the generated data. A low consistent value distributed across all subject is best as it indicates the subject-biometric classification network is unable to find discriminative features. 144
6.12 The distribution of data points embedding of subject invariant signals coloured by frequency. 145

A.1 Main interface of the Cognionics Data Acquisition software 184
A.2 Pre-experimentation configuration including sensor selection 184
A.3 Real-time signals streaming with real-time impedance value for each sensor 185

B.1 The coordinate format for window display for both software, where (0, 0) is the origin and (X, Y) is the width and height of the monitor resolution used. 188

List of Tables

2.1 SSVEP-based application for speller applications 21
2.2 SSVEP-based application for robotic applications 23

3.1 Frame arrays set for the frequency modulations 48
3.2 Definition and Notation use in Algorthm 1 and Algorithm 2 49
3.3 Parameters used in the CNN Model 57
3.4 Parameters used in the RNN Model 59
3.5 Mean accuracy with standard deviation over 10-fold cross validation for subject, S01 60
3.6 Mean accuracy with standard deviation over 10-fold cross validation for each of the three subjects, mean results across subjects are also presented 62
3.7 Test Accuracy across subjects 62

4.1 Mean accuracy and ITR with standard deviation for calibration classification over 10-fold cross validation. 81
4.2 Accuracy for each experiment and mean accuracy and ITR with standard deviation for real-time classification. 84
4.3 Mean accuracy with standard deviation over 10-fold cross validation for nine dry-EEG subjects. 89
4.4 Test Accuracy across subjects 89
4.5 Mean accuracy with standard deviation when classifying SSVEP for unseen subject by leave-one-subject-out. 90

5.1 Classification test accuracy using generated and real-world data used to train the classifier. The baseline result is the classification of only real data. 110

5.2 Accuracy for test classification using synthetic data for pre-training stage. The baseline contains no pre-training. 112
5.3 Classification test accuracy for cross-subject Generalisation (see Figure 5.12) via pre-training using synthetic data. The baseline does not include a pre-training stage. 115

6.1 Mean accuracy with standard deviation when classifying SSVEP for unseen subject on offline dataset. 137
6.2 Mean accuracy with standard deviation when classifying SSVEP for unseen subject on offline dataset. 140
6.3 Mean accuracy with standard deviation for cross-task when classifying SSVEP on online subject dataset. 141

List of Algorithms

1 Window flipping 48

2 Streaming bio-signals from Cognionics data acquisition 185

3 Displaying variable stimuli with flickering bounding boxes on psychopy via SSD . . . 188

List of Abbreviations

AC-GAN Auxiliary Classifier GAN.

ALS Amyotrophic Lateral Sclerosis.

BCI Brain-Computer Interface.

CCA Canonical Correlation Analysis.

cDCGAN conditional Deep Convolutional GAN.

CNN Convolutional Neural Network.

CSP Common Spatial Pattern.

DC-GAN Deep Convolutional Generative Adversarial Network.

DNN Deep Neural networks.

EEG Electroencephalography.

ERP Event-Related Potential.

FairGAN Fairness-aware Generative Adversarial Network.

FFT Fast Fourier Transform.

GAN Generative Adversarial Networks.

GP Gradient Penalty.

GRU Gated Recurrent Units.

HCI Human-Computer Interaction.

HR High-resolution.

ICA Independent Component Analysis.

ITR Information Transfer Rate.

JFPM Joint Frequency-Phase Modulation.

k-NN k-Nearest Neighbour.

KF Kalman Filter.

LDA Linear Discriminant Analysis.

LIS Locked-In Syndrome.

LR Low-resolution.

LSTM Long-short Term Memory.

MDM Minimum Distance to Mean.

MI Motor Imagery.

ML Machine learning.

MLP Multi-layer Perceptron.

MSI Multivariate Synchronization Index.

MTF Markov Transition Field.

PReLU Parametric Rectified Linear Unit.

PSD Power Spectral Density.

RDA Remote Data Access.

ReLU Rectified Linear Unit.

RGAN Recurrent GAN.

RMS Root Mean Square.

RNN Recurrent Neural Network.

SCU SSVEP Convolutional Unit.

SIS-GAN Subject Invariant SSVEP Generative Adversarial Network.

SNR Signal-to-Noise Ratio.

SR Super-resolution.

SSVEP Steady-State Visual Evoked Potential.

STM Style Transfer Mapping.

SVM Support Vector Machine.

SWLDA Stepwise Linear Discriminant Analysis.

t-SNE T-distributed Stochastic Neighbor Embedding.

TFR Time-frequency Representation.

VAE Variational Auto-encoders.

WGAN Wasserstein GAN.

Chapter 1

Introduction

This thesis explores novel Human-Computer Interaction (HCI) via the use of Brain-Computer Interface (BCI) and the subsequent use of recent advances in deep learning for the effective decoding of cortical signals in real-time.

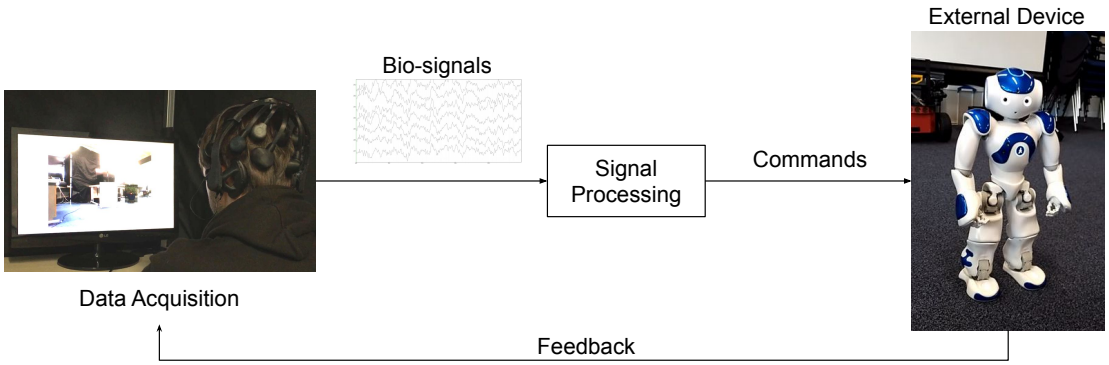


Figure 1.1: Block diagram of the main BCI components

The field of HCI encompasses a range of topics which study how best to facilitate communication between humans and machines. Communication is an essential element for humans in daily life, either between groups of people or with technologies that now form a crucial part of our everyday life. Globally, there are people who suffer from neuromuscular diseases and who have

severe disabilities such as complete Locked-In Syndrome (LIS) or Amyotrophic Lateral Sclerosis (ALS), which limits their fundamental communicative ability.

In such cases, communication can be partially revived by utilising brainwaves measured using an acquisition device. By exploring any intention contained within these captured bio-signals, they can be translated and exploited as certain commands, thus providing a potential way of communication [64, 193, 236]. We call such a communication medium Brain-Computer Interface. Figure 1.1 shows the main components of a BCI system, where the subject (healthy participant or patient) wears a device that is able to measure cortical signals from the scalp. These bio-signals represent cortical neural ensemble responses from a certain mental task, either voluntarily or evoked by external stimuli. These signals can be interpreted as commands for any device by decoding them using signal processing techniques. Accurate decoding can then be utilised to assist humans in any way required, either virtually or physically.

Researchers originally delved into BCI over half a century ago [113], and since then it has been actively developed. However, there has thus far been a distinct lack of applicability of such approaches outside of the research laboratory [30, 160]. This impediment can essentially be attributed to the cumbersome nature of the equipment, the rigid experimentation protocols required and the challenges of accurate signal decoding, especially in making a system work in real-time.

This thesis focuses on tackling these constraints in order to provide fundamental enablers in making BCI systems more practically applicable for use in assistive technologies within everyday life. These constraints are discussed in detail in the next section as the primary research questions of this thesis, as well as presenting the motivation for tackling them.

1.1 Research Questions and Motivation

Continuously growing enthusiasm in BCI research has been a positive development towards making a BCI system applicable and ultimately to achieve the primary goal of the field - to provide an accurate, comfortable and applicable non-muscular communication solution. However, some constraints prevent the advancement of such technology for use outside of a research environment.

The signal acquisition procedure is the most crucial part in BCI as it forms the input to the system. This stage involves the data acquisition device to measure the bio-signals, the interface to record the signals, the stimuli design to evoke the neural responses, and the experimentation protocol. The bio-signals gathered in this stage are the input to the BCI system as highlighted in Figure 1.1. Hence, to ensure the system has gathered satisfactory bio-signals, researchers apply a rigid measuring procedure covering the device used, the stimuli design and to the whole protocol – ultimately resulting in a system that is difficult to be used daily.

Due to the non-invasive nature, Electroencephalography (EEG) is a favoured technique among BCI researchers. However, it often features a complicated usage process that currently seems possible only in a laboratory [160]. Typically, EEG requires pre-and post-skin preparation protocol for the application of electrode gel, owing to the Ag/AgCl (silver/silver chloride) electrode that is not only time consuming but also inconvenient to apply. Other traits of EEG such as wired devices and the requirement of a physical shield establish the rigid experimental protocols.

With recent progress in technology, alternative devices use dry sensors technology makes it possible to eliminate such a cumbersome protocol and make it viable for use as frequently as needed, even in a real-time application. However, this alternative solution might jeopardise the quality of the signals, resulting in a substantially more challenging signal decoding and classification task, thus possibly giving out an inaccurate translation [154]. In favour of compensating for this shortcoming, the system needs to have a favourable signal processing approach that is capable of decoding BCI signals whilst similarly enabling the use of this less cumbersome dry-sensor technology.

Brain activity can be induced by an external stimulus that makes it possible to evoke a neuro-physiological response. Steady-State Visual Evoked Potential (SSVEP) is a type of neuro-physiological response induced directly via subject fixation on a frequency-based visual stimuli and requires no *a priori* user training [137, 45, 169]. The frequencies used as stimuli are determined by a few factors such as display device specification, the number of frequencies required and harmonics frequencies which limits the number of frequencies that can be used [177, 247]. Commonly, these visual stimuli are known for having similar size stimulus boxes with a fixed position and similar pixel region within an application which does not reflect user interaction with a natural scene or environment.

While it is essential to have access to large quantities of data for training a machine learning based signal decoding method, collecting high-quality EEG data has proven difficult [253, 145]. The experimental procedure is time-consuming, expensive and difficult to operate within the usually short amount of time experimental subjects can perform EEG experiments [53]. In addition to these issues, the pattern of electrical voltage produced by the brain often differs from one subject to another, making BCI is a highly subject variant [141, 56, 235, 117, 248].

Commonly, because of various external factors, the EEG signal patterns for a given individual can appear differently [10], forcing per-session calibration for any given subject for effective utilisation. The reasons can be, but are not limited to, the diverse nature of subject intra-session, the condition of the acquisition device and the signal quality decay over the session duration [112, 2, 250]. This drawback leads to a long calibration process for real-time applications, restraining the practical application of EEG systems. The decoding process associates similar attributes in signals with a label for accurate prediction; thus, the need for the calibration stage to ensure having consistent signals.

Based on this range of constraints, this thesis focuses on addressing the following primary research questions:

- Can the cumbersome signal acquisition procedure be eliminated without jeopardising the signal decoding process?
- Can a real-time BCI application be created within a limited frequency range, whilst also using a natural environment based representation as the basis for the stimuli?
- How can we best to address the problem of data collection, both in term of the procedure and, subject and session variance which currently results in small datasets and an expensive calibration stage?

Deep Neural Networks (DNN) are a state-of-the-art machine learning approach that has the ability to discriminate data based on their characteristic features. DNN approaches are more suitable for noisier signals than traditional machine learning approaches, as they eliminate the need for bespoke feature extraction and representation processes. DNN are able to automatically learn the best representation of data by adjusting parameters based on error computed against some training objective [81].

Commonly in the computer vision field, a task such as an object detection or scene recognition is accomplished using DNN [121]. Taking advantage of such advances, it is possible to use DNN to detect objects in a given scene in real-time, and for this to be implemented as the basis of SSVEP stimuli. The frequency range limitation in SSVEP stimuli can possibly be tackled by using a multi-layered stimuli, where each frequency is assigned to different objects that appear in a frame at a time. To design stimuli that appear more natural to subjects, the pixel region for the stimulus can be identified using an object detection algorithm within the real-world natural environment derived from the incoming video stream from a teleoperated robot camera. Detected objects can be displayed as stimuli, where both the dimensions of the flicker regions and their on-screen position changes depending on the scene objects detected.

The versatility of DNN to imitate real bio-signals by learning their characteristic attributes draws the possibility for the algorithm to generate sets of synthetic data [82]. This can be utilised by training the generative model to learn specific features and to potentially unlearn any biased features. This may create realistic bio-signals without biased features, maybe enabling such approaches to perform better in classifying unseen subjects.

Overall, this thesis focuses on exploring the use of recent advances in machine learning, notably deep architectures, to tackle certain critical limitations within the area of SSVEP-based BCI, such as rigid experimentation procedure, limited frequency ranges, hand-crafted classifier, small dataset, expensive calibration stage. Three potential deep learning techniques to be explored will be bio-signal pattern recognition, visual object detection for stimuli creation and synthetic bio-signal generation.

1.2 Research Aim and Objectives

The primary aim of this thesis is to tackle some of the fundamental constraints in BCI applications (as detailed in Section 1.1), whilst focusing on the SSVEP paradigm and exploiting the advancement of Deep Neural Networks. In order to achieve this primary aim, the following research objectives are set:

- Investigate the ability of the commercial dry-sensor technology to provide a less rigid experimentation procedure without jeopardising the performance of the signal prediction by evaluating classification accuracy.

- Incorporate state-of-the-art computer vision techniques for object detection in designing real-life representation stimuli for real-time teleoperation BCI application by evaluating the information transfer rate and classification accuracy.
- Devise and evaluate the optimal approaches within state-of-the-art generative models for eliminating the calibration stage in a BCI application, by tackling both session and subject variant problems by evaluating classification accuracy on unseen subjects.

These research objectives are going to be evaluated differently based on the objective:

- The first objective will be evaluated by measuring the accuracy of the SSVEP classification on the bio-signals collected by using the proposed method to ensure that the procedure has been done scientifically correct. We also measure the accuracy performance of the proposed DNN classifier against traditional machine learning methods.
- We evaluate the second objective by assessing if the proposed real-life representation stimuli works by measuring the accuracy and information transfer rate during a real-time experiment comparing to the state-of-the-art application.
- The last objective is evaluated by verifying the ability of the proposed generative model in tackling subject variance by comparing the accuracy performance when classifying SSVEP bio-signals of unseen subject, when using the collected bio-signals, mixing the generated signals with the real bio-signals and when just using the generated signals alone.

1.3 Thesis Contribution

This thesis presents contributions to the state-of-the-art in the following areas:

- The development of an improved method for a SSVEP-based dry-EEG system in place of previous state-of-the-art approaches [115, 226, 120] via a deep neural network, eliminating the need for hand-crafted manual feature extraction as compared to prior work [133, 31, 135].

- Creation of a novel variable position and size SSVEP stimuli, using objects detected in real-time from the natural environment. The stimuli is generated within the live video stream from a teleoperated humanoid robot traversing a real-world natural environment. This has notable differences from previous work [58, 257, 64, 137] in terms of stimuli pixel pattern, size and on-screen position, and uniquely presents the user with a natural environment interface for stimuli driven BCI.
- An exploration of synthetic EEG-data generation containing SSVEP signals via a variety of unsupervised models. The techniques generate realistic synthetic signals directly in the same domain for classification task improvement, unlike prior works [253, 191].
- A novel neural-based generative model, creating highly-realistic and enforced subject invariant synthetic EEG data for downstream predictive tasks. The generated subject-invariant bio-signals are capable of classifying unseen subjects in comparison to the prior work [112, 98], providing the potential of eliminating the calibration stage for real-time BCI applications.

1.4 Publications

To date, research presented in this thesis has been published in the following peer-reviewed articles:

- Nik Khadijah Nik Aznan, Stephen Bonner, Jason Connolly, Noura Al Moubayed, and Toby P Breckon. On the Classification of SSVEP-based dry-EEG Signals via Convolutional Neural Networks. In *International Conference on Systems, Man, and Cybernetics*, pages 3726–3731. IEEE, 2018
- Nik Khadijah Nik Aznan, Jason D Connolly, Noura Al Moubayed, and Toby P Breckon. Using Variable Natural Environment Brain-Computer Interface Stimuli for Real-Time Humanoid Robot Navigation. In *International Conference Robotics and Automation*, pages 4889–4895. IEEE, 2019
- Nik Khadijah Nik Aznan, Amir Atapour-Abarghouei, Stephen Bonner, Jason D Connolly, Noura Al Moubayed, and Toby P Breckon. Simulating Brain Signals: Creating Synthetic EEG Data via Neural-based Generative Models for Improved SSVEP Classification. In *International Joint Conference on Neural Networks*, pages 1–8. IEEE, 2019

- Nik Khadijah Nik Aznan, Amir Atapour-Abarghouei, Stephen Bonner, Jason D Connolly, and Toby P Breckon. Leveraging Synthetic Subject Invariant EEG Signals for Zero Calibration BCI. In *International Conference on Pattern Recognition*. IEEE, 2020

1.5 Scope

In this thesis, we investigate the adequacy of DNN in undertaking various limitations in the field of Brain-Computer Interface. It is worth noting that the work in this thesis is limited by several factors, such as the availability of dataset and hardware.

The bio-signals used in every chapter of the thesis are collected using dry-EEG technology as the data acquisition hardware and SSVEP as the frequency-based stimuli. We cover different machine learning approaches for bio-signal evaluation to ensure the quality of bio-signals are pertinent and an attestation of optimal signal decoding. However, this research does not include any data collection using conventional wet-EEG technology; thus, no side-by-side comparison between the two is presented. Although there are a small number of available open-source datasets containing SSVEP-based bio-signals, due to the different frequency classes and application settings used, no other dataset has been used in any part of this thesis except from our collected dataset.

The data acquisition procedures that have been used for this thesis have been evaluated in real-time experiments. It is a crucial step within a BCI system to prove the ability of the system to provide communication in a real life setting. Although in this thesis we have successfully conducted a teleoperation BCI application under real-time operating conditions, due to the limitation of the on-board camera of the humanoid robot, the experiments have been conducted under controlled settings. This is due to missing depth information of the objects, meaning that the application must obtain the object heights for the calculation of the distance and the direction of the robot trajectory.

All the bio-signals collected during the period of this work were taken from experiments conducted within departments in Durham University[1] where all of our experiments and evaluations

[1] The experiments conducted for this thesis have been approved by the Ethics Sub-committee in the Department of Computer Science and Department of Psychology from Durham University, United Kingdom. All subjects are informed that their data is collected and have signed the consent form.

have only been performed on healthy subjects[2]. Hence no real patients suffering from a medical condition have been used as subjects in this thesis.

1.6 Thesis Structure

The background of the main constituents of this research are introduced in **Chapter 2**. We also present a review of the literature related to every component consistent with the contributions of this thesis. The review is focused on BCI elements which are the signal acquisition, neuro-physiological response and bio-signal pattern recognition. Furthermore, this chapter highlights the limitations of the current state-of-the-art in BCI that is relevant to this present work.

In **Chapter 3**, we present the details of the BCI experimentation protocol, such as the application of the dry-EEG device, the stimuli used and the process of data collection. We focuses on integrating a dry-EEG headset with SSVEP stimuli to demonstrate the appropriate of bio-signals collected. The performance of bio-signals are evaluated using several approaches from both traditional machine learning and deep neural networks in order to show the ability of our proposed technique against others.

We move forward in **Chapter 4** by integrating a real-time application to navigate a humanoid robot in a natural environment. We leverage DNN based image and signal understanding to facilitate both real-time object detection and bio-signal decoding. We employ the detected objects acquired by the humanoid robot from the environment currently being navigated as the novel variable SSVEP stimuli. On-screen object selection from subjects facilitates the online decoding of human cortical brain signals directly into teleoperation robot commands.

Chapter 5 explores the possibility of generating synthetic bio-signals using a variety of unsupervised neural network. We investigate the potential of synthetic data in capturing the character of SSVEP bio-signals in order to tackle the shortage of reliable data and in improving the performance of classifying real-world bio-signals.

Taking into consideration the biased features learned by a neural network that limits the ability to uphold the performance in a new subject dataset, in **Chapter 6**, a subject invariant

[2] Healthy subjects are those that do not suffer from LIS or ALS conditions and self-declared healthy conditions in general by each participant.

approach is explored. We investigate the potential of this approach in generating highly-realistic synthetic EEG data invariant to any subject, session or other environmental conditions. We demonstrate unseen subject evaluation in both the supervised model and unsupervised model that includes the potential of eliminating the calibration stage for real-time BCI applications.

Finally, a summary of contributions and results present in this thesis is detailed in **Chapter 7**. Additionally, some limitations and future direction aligned to this thesis are also discussed.

Chapter 2

Literature Review

In this chapter, we review current state-of-the-art from the fields of BCI and machine learning. Specifically, we focus on current research that is exploring the use of a dry-EEG system for signal acquisition to find an alternative way of gathering bio-signals that are able to reduce the rigid procedure and the preparation time of the experiment. In terms of the neuro-physiological response, we focus on current work that implements SSVEP stimuli for the system to see the potential way of improving the stimuli display for a real-time natural environment application. We also investigate a deep neural network as the prediction method on the bio-signals especially on dry-EEG bio-signals. Additionally, we review if deep neural networks can be used in EEG data generation to potentially move towards zero-calibration application, tackling the subject variant problem. Moreover, we draw attention to some limitations and gaps within literature and pertinence to the present thesis.

2.1 BCI Signal Acquisition

A Brain-Computer Interface (BCI) is a system that provides a communication and control medium between human cortical signals and external devices [194, 239, 31, 166, 5, 156, 142]. There are two primary acquisition methods to measure brain activity:- invasive, which involves sensor implantation in the brain and non-invasive, which uses external sensors [2, 3].

This thesis will focus on non-invasive data acquisition approaches, specifically on Electroencephalography (EEG) as it is the most suitable approach within BCI research development since it has the ability to measure the cortical waves without the need for implanting any sensor directly into the brain.

2.1.1 EEG-based BCI

EEG is a cortical acquisition technique where electrodes or sensors are placed on the scalp to capture the electrical activity of the brain. It is the most notable signal acquisition technique in BCI, especially for scientific purposes as it is a non-invasive process of measuring and recording cortical activity [148, 216, 30, 229, 193]. The sensor placement typically follows the standard 10-20 system, instead of implanting electrodes directly in the brain like invasive signal acquisition, EEG measures the cortical signal via electrodes or sensors placed on the surface of the scalp [230, 2, 5, 3].

However, this results in poor spatial resolution signals with a low signal-to-noise ratio (SNR) and interference from a signal acquisition standpoint by other muscle activities such as eye blinks or head movements [160, 201]. Despite such limitations, EEG is safe and easy to use, relatively affordable and has an excellent temporal resolution, thus making EEG the most dominant BCI signal acquisition technique to date [123, 197, 3, 152].

Conventionally, the majority of EEG acquisition devices rely on the cumbersome application of conductive gel or paste (so-called wet-EEG) to reduce the impedance between the Ag/AgCl electrode and the scalp [139, 148, 160, 243]. The impedance values in EEG signals are a measurement of how good the conductivity is between the electrode and the skin. The lower the value of impedance, the better the electrode and the skin contact thus improving overall EEG signal quality [55, 165, 69, 154].

The required gel application is a significant drawback of this technology where it is used to fill in the gap between electrodes and scalp every time before using the traditional EEG cap [139, 165, 105]. This consequently results in relatively substantial preparation time and also includes scratching of the scalp and additional time to remove the gel after the experimental protocol has finished [195, 110, 239, 249, 182]. This contributes to a lengthy, time consuming and often uncomfortable experience for the participants. Furthermore, the gel will dry over a

specific time frame and increase the impedance value, thus somewhat limiting the experimental data acquisition interval [139, 195, 73, 156].

Moreover, classical wet-EEG requires some specific experimental conditions like a Faraday cage (a physical shield using conductive material) which reduces the effect of external electromagnetic interference in terms of signal noise [190, 160]. The bulky size and wired connectivity requirements of the wet-EEG also limits the use of the device in different kinds of environments [130, 45], thus limiting the application of BCI using wet-EEG to strict experimental operating conditions.

(a) An example of wet EEG electrode cap from Biosemi Active Two System [17].

(b) An example of dry EEG headset from Cognionics Inc.

Figure 2.1: Example of wet EEG technology and dry EEG technology

There is a relatively vast number of studies on dry-EEG as an alternative approach alleviates these limitations in terms of skin preparation, stable connectivity and comfort during experimentation [135, 154, 161, 242] in addition to ready adaptability to different head sizes [163, 133]. A Dry-EEG system uses dry electrodes that perform the same function as wet electrodes but instead of using electrode gel to create connectivity between electrodes and the scalp, dry electrodes are designed to be able to slide through the hair and create sufficient force for direct contact against

the scalp [163, 110, 154].

2.1.2 Dry-EEG vs Wet-EEG System

In striving for a better EEG system, there are many technologies and research works on EEG systems with different approaches for replacing the traditional wet-EEG. The authors of [55] compare four EEG systems; one traditional wired wet-EEG (ActiveTwo), two wireless systems (B-Alert X10, a gel-infused foam system and EPOC, a saline-infused felt system) and one dry-EEG system (Quasar dry sensor interface). The saline-based EEG system works like wet-EEG where the sensors require the use of saline solutions as a connection bridge to the scalp therefore requires pre-experiment procedure. Unlike wet-EEG that requires gel-based solution, saline-based system is easier in terms of post-experiment procedure. In this work, the authors conduct an identical experiment four times repetitively, one for each device on 16 subjects performing seven cognitive tasks. The subjects evaluated each device by rating them on different aspects such as; different head sizes adaptability, subject preference and comfortability (pressure on the scalp from the device, the weight of the device and application procedure), scalp location availability (consistency between application of a subject and availability location for particular signal of interest and average), electrode connection stability and experimental integration setup (data recording method). This study suggest that, there is room for improvement for a wireless dry-EEG system to be the most prominent system in BCI.

Similarly, in [186], the authors evaluate three different electrodes - two wet EEG systems (hydrogel-based and water-based) and a dry electrode-based EEG system. They use g.GAMMAsys from Guger Technologies (g.tec) with g.LADYbird active wet electrodes. The water-based electrode system is from TMSi, a wireless passive water-based wet-EEG system. Even though the system uses wet electrodes, they eliminate the necessity of post-protocol such as hair washing as only water is involved. The dry-EEG system used in this paper is from g.tec as well, (g.Sahara) where the electrodes are dry electrodes made from gold alloy pin to better penetrate through hair to make a connection with the scalp. The paper evaluates each system based on a technical and user-centred test on eight healthy subjects. The technical test simulates short circuit noise produced by each electrode while attached to a polished copper plate for 12 minutes. The dry electrode-based system has the highest mean RMS (Root Mean Square) noise value, and the water-based system has the lowest value. The user-centred test consists of two parts; a P300 task and questionnaires for subjects. Subjects were required to complete four P300 communication

and control tasks; spelling two words, controlling a multimedia player, navigating a web browser and spelling another two words. Before the actual task, each subject was required to complete a calibration training stage where their cortical signals were recorded, and they needed to spell a word using a P300 matrix speller. The recorded signals were used to train an SWLDA (Stepwise Linear Discriminant Analysis) classifier for the prediction of the actual task. Based on all the assessments, on average the hydrogel-based system gave the highest classification accuracy, whilst the dry-EEG system gave the lowest scores. From the questionnaire results, taking the mean across questionnaires and subjects, the water-based get the highest score followed by the hydrogel-based and dry electrode-based system.

A comparison study presented in [154], compares three different EEG systems from Brain Products. These are: wet electrodes EEG without amplifier (passive wet-EEG), wet electrodes EEG with an amplifier (active wet-EEG) and dry electrodes EEG with an amplifier (active dry-EEG). Three identical experiments were repeated for all EEG types on eight healthy subjects completing an auditory task. The task was an ERP (Event-Related Potential) neuro-physiological response where it measures P3 response evoked by audio stimuli. Subjects were required to sit still with their right hand rested on the table in front of them whilst having a fixation on a white cross on a black screen in front of them. They needed to then press the space bar on a keyboard using their right-hand finger when they heard any rare tone. The authors compare the cortical signals collected from the experiment in term of EEG spectra by plotting the raw data transformed via a Fast Fourier Transform (FFT), single-trial noise by computing RMS, ERP analysis by analysing the error level of the P3 responses and ERP power by computing the number of samples required for reaching significant statistical power. The results show that dry-EEG performs the worst when compared to two other systems with high noise and impedance level, which contribute to weak quality signals. However, this study proves the ability of dry electrodes to measure comparable cortical signals with significant advantages of easy to use, thus with proper techniques and better dry electrodes development; they are promising.

A simple resting-state EEG experiment was conducted in [152] to compare the quality of cortical signals from the Cognionics Quick-20 dry-EEG system to the Brain Products actiCAP wet-EEG system. The dry-EEG headset is a fully dry system with Bluetooth connectivity and portable while the wet-EEG is fully wired with active wet electrodes. The data was collected from 12 subjects in two sessions, one for each system. Subjects sat on a chair in a dim room for resting state where they rest with opened eyes for five minutes and rest with closed eyes for another five minutes while having an EEG system recording the cortical signals. The authors

use two evaluation metrics to compare both systems; Power Spectral Density (PSD) and alpha suppression. Pearson correlation coefficient is computed to find the correlation of log PSD across systems as similarity measurement between electrodes channels. The results show that both systems are highly correlated on average on every channel and every subject for both tasks (opened eyes and closed eyes), although dry-EEG system produces more noise than the wet-EEG system at low frequencies. Alpha suppression is a measurement of the reduction in alpha-band power (8-12Hz) from opened eyes to closed eyes. On average, across subjects, both systems show alpha suppression with alpha-band power is greater during closed eyes rather than opened eyes. From this comparison study, the dry-EEG system is shown to be comparable to the traditional wet-EEG system in terms of data quality during resting-state activity.

From all the available literature comparing different types of EEG system, it can be seen that a dry-EEG system is largely comparable with a wet-EEG system. However, dry-EEG suffers from inherently reduced signal quality due to the lack of conduit gel and the elimination of rigid procedure such as active amplifier and Faraday cage, Dry-EGG requires a better approach to process and extract knowledge from the collected signal.

2.1.3 Cognionics Dry-EEG System

There are several commercial EEG systems in the market designed to improve the subject experience in terms of comfort and overall application experiences. This thesis will focus on reviewing BCI-based applications using data collected by dry-EEG technology from Cognionics Inc and investigate how well it performs at different paradigms and applications when using different prediction methods. This section will review the Cognionics dry-EEG system, as well as research studies using it for various applications.

Cognionics offers a range of different systems depending upon the number of channels, applications and price. All their dry-EEG systems use a fully dry-electrodes comb or claw-looking sensor that is designed to flex through the hair with a flexible angled leg that is able to bend to give mechanics pressure on the scalp for functional connectivity [44, 162, 31]. These electrodes remove the need for conventional experiment protocol such as skin preparation, electrode gel filling and post-experiment clean-up [133, 135, 182, 152]. All the systems are a wireless, portable and have built-in Faraday cages in the individual electrodes and feature an adjustable headset

for different head sizes to make them more practical for different subjects and applications [44, 163, 31].

The work in [132] takes advantage of the wireless feature of the 32-channel Cognionics system to explore the quality of bio-signals for the SSVEP task whilst walking as compared to the standing position. The experiment required ten healthy subjects to take turns walking on a treadmill with four different speeds; 0 (standing position), 1, 2 and 3 miles per hour while having a fixation on the SSVEP stimuli with eight dry electrodes to record the cortical signals. They evaluated the performance of the signals by applying Canonical Correlation Analysis (CCA) and Independent Component Analysis (ICA) to classify the SSVEP signals. As expected, the faster the speed, the more inferior the quality of the signals. However, this literature opens the possibility of real-life applications that are not limited to laboratory use.

Later the authors repeated the same task in order to compare the performance of dry-EEG to wet-EEG during walking trials [133]. They employed the same system as their previous work with two additional disposable wet electrodes for wet-EEG signal recording. A similar experimental protocol was conducted on 19 healthy subjects, and from their experiments, wet-EEG performed better as compared to dry-EEG by 4% to 10% accuracy for standing and walking at different speeds, respectively.

Inspired by the previous study that highlighted the ability to record EEG signals in motion, the authors in [31] used a 64 channel Cognionics system for real-world aviation applications. These involved using a flight simulator motion platform and also flying a real open cockpit biplane. EEG data was recorded as the pilot flew whilst an auditory stimulus event was played in the background. The goal being that using the EEG signal, it should be possible to identify when an audio stimulus is present during both experimental conditions. The authors proposed the use of ICA combined with a Kalman Filter (KF) for classifying the signals. One of the main aims of this paper is to show that a dry-EEG system is capable of recording good quality signals outside the laboratory and under extremely challenging conditions, thus adding further evidence that dry-EEG-based BCI is applicable in the real-world.

Another study using using a 20-channel Cognionics headset for foot motor imagery was carried out in [135] to trigger a lower limb exoskeleton robot (XoR). The paper aimed to have a quick setup system for asynchronous motor imagery BCI as offered by using the dry-EEG headset. EEG data was collected from three healthy subjects for a calibration stage where each subject

was required to perform four runs of ten trials of foot imagery task where they were given a cue to execute the motor imagery. The recorded data was used to decode the signals for threshold tuning for a resting class and a foot imagery class and to execute XoR during the online task. By using Common Spatial Pattern (CSP) and sliding windows for feature extraction, they trained a logistic regression to compute the probability between zero and one to tune the squatting execution threshold. During the online experiment, subjects were required to perform foot imagery on a given cue, and, if the signals reached the foot imagery class threshold, XoR was triggered. Two of the three subjects exceed 70% test accuracy during an online experiment using this approach.

Overall, it can be seen that although a dry-EEG system is comparable to the wet-EEG system, dry-EEG bio-signals have relatively higher noise levels and higher impedance values when compared to wet-EEG [133, 55, 186, 154, 152, 146], thus resulting in a substantially more challenging signal decoding and classification tasks.

In this thesis, taking the advantages of the dry-EEG system in making BCI application more applicable and practical, we tackle and explore different techniques aimed at improving bio-signals decoding performance. The technicalities of the experimentation procedure of the dry-EEG system are in Chapter 3 of this thesis.

2.2 BCI Neuro-physiological Responses

There are different types of neuro-physiological responses that can be used for BCI application based on the type of cortical neural ensemble responses [5, 243]. These responses can be grouped into two primary types: endogenous and exogenous [194, 5]. An endogenous approach is when a subject self initiates any control voluntarily without relying on stimuli, while the exogenous type is when there are external stimuli involved to trigger the control [62, 174, 90, 3].

2.2.1 Type of Responses

Motor Imagery (MI) is a response that does not depend on any external stimuli where the primary motor cortex becomes active without any overt muscle movement (imagination only)

[5, 160, 3]. This imagination process produces similar neural activity, although with smaller magnitude to the actual movement [194, 151, 146]. This response requires a very long training time for the imagined movement to become detectable via the hardware [70, 156, 233]. This training stage is needed to familiarise the participant with the EEG experiment and the process of imagining movement [161]. Apart from the substantial training stage, another drawback is that only a maximum of four control tasks are suitable for MI, in order to ensure excellent quality signals [143].

Stimulus-evoked is a type of response with almost no training time needed for the participant as the brain produces a response towards a stimulus [31]. Within this type, a common example is the P300 potential where the cortical signals respond with a positive deflection after 300ms via the use of a salient external sensory stimulus [148, 151, 146]. P300 is named as such based on its polarity, which is the positive deflection in voltage and fixed latency (delay between stimulus and response) of roughly 250 to 500 ms [26, 186, 156, 103]. It is also known as an oddball paradigm as the quality of cortical signals depend on the subject attention on the rare occurrence of targeted stimuli; therefore, the quality might decrease after some time as the stimuli become more familiar or subject pays less attention to it [153, 165, 169]. P300 has a long delay between stimuli [174] such that it decreases the effective Information Transfer Rate (ITR) achievable by the system. ITR measures the accuracy and the speed of a BCI system in real-time by showing the total amount of information that is successfully transferred from the EEG device to the application per time unit [216].

Steady-State Visual Evoked Potential (SSVEP) is another type of stimulus-evoked response which is induced simply via subject fixation (or even just via peripheral attention) on periodic visual stimuli and requires almost no *a priori* user training [137, 45, 169, 102]. On visual stimuli, the subject is presented with repeated alternating white and black frames flashing at specific frequencies which induces the primary visual areas to oscillate with the same frequency and harmonics [7, 256, 151, 103, 247]. SSVEP has the feature of frequency tagging, which enables the measurement of neural activity in response to a flickering stimuli which the subject is fixated upon, even if the subject is not paying full conscious attention to the stimuli [174, 33, 203]. It is considered to be the most suitable type of stimuli for effective high throughput BCI as SSVEP signals are predominant and can be detected relativity simply. As a result, they provide high ITR neural signals with minimal subject training [153, 166, 115, 103, 255].

A comparison study between these two stimulus-evoked responses has been performed in [52], where authors compare the performance of a BCI speller between P300 (or P3)-based BCI

and SSVEP-based BCI on seven patients suffering from Lock in Syndrome (LIS). The cortical signals were recorded using a wet-EEG system for both responses in different experiments for evaluating the responses in term of accuracy, ITR performance and subjects preference. This work is divided into two main sessions for both BCI paradigms, which are preliminary session and acquisition session. A preliminary session for each response is conducted on every subject before the acquisition session, which familiarises the system to subjects and optimises the paradigms by setting up the stimulus based on individual preferences. During the P300 preliminary session, the stimulus setup is tailored based on stimulation style (row/column or single symbol), stimulus duration, interstimulus interval and flash sequence duration. Whilst during the SSVEP session, the optimum stimulation frequencies are chosen. The results from these sessions are not used for analysis purposes. During the P300 acquisition session, subjects sat in front of a computer screen with a six-by-six alphanumeric matrix. Subjects count how many times a symbol was highlighted to choose the symbol. The calibration phase was performed at the beginning of the acquisition session to build the classifier for decoding the signals during the actual session. During the actual session, this classifier was used to predict the words spelled by subjects. For the SSVEP speller, the subjects sat in front of a 60Hz LCD screen with 4 boxes. All boxes flashed with a different frequency, with frequency ranges between 6.67 Hz to 15 Hz depended upon subjects preference in the preliminary session. Subjects spell words by choosing one of the boxes at a time, where each box contains 16 symbols at the start. After choosing a box, the symbols in the box are divided into 4 other boxes to be chosen by subjects. This process was repeated until a symbol is chosen, and ultimately a word is spelt. Based on the performance result, SSVEP speller outperformed the P300 speller in term of accuracy, speed and ITR.

In this thesis, taking advantage of the effectiveness compared with competing approaches, we utilise SSVEP as the neuro-physiological paradigm for data collection. There are many vital applications made possible in BCI via the use of SSVEP. For example, it can be utilised to allow people with severe physical disabilities to control or communicate with external devices just by having them fixate on flickering stimuli [56], controlling an exoskeleton [115] or navigating a humanoid robot [218].

2.2.2 SSVEP-based BCI

There numerous different EEG-based BCI application using SSVEP as a neuro-physiological response. This section discusses different applications using SSVEP, the design of the stimuli, common frequencies used and other details regarding the stimuli.

SSVEP-based Speller

One of the useful and prevalent applications in EEG-based BCI is the 'speller', where instead of using motor muscle movement to type a word, subjects are able to 'type' by looking at SSVEP stimuli to trigger the visual cortex [197]. This is achieved by assigning each letter to a unique SSVEP frequency. The literature on SSVEP-based spellers encompasses a wide range of choices on frequency, speller design and display device. A selection of approaches is highlighted in Table 2.1.

Experiment Type	Frequency (Hz)	Display device
40 alphanumeric stimuli [40]	[8.0, 15.8] step of 0.2Hz with phase: [0.0 π, 1.90π).	60Hz refresh rate LCD monitor.
28 alphabet stimuli [24]	[6.10, 11.28]	120Hz refresh rate Laptop
12 numerical stimuli [166, 235]	[9.25, 14.75] step of 0.5Hz with phase: [0.0 π, 1.5π] step of 0.5π	60Hz refresh rate LCD monitor
Multi-layer 30 alphabet stimuli [238]	{13, 14, 15} with phase: {0rad, πrad}	LCD monitor
Multi-layer 58 alphanumeric stimuli [238]	{6.67, 7.50, 8.57, 10.00, 12.00}	60Hz refresh rate LCD monitor
Multi-layer 27 alphabet stimuli [238]	{6.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0}	Web browser

Table 2.1: SSVEP-based application for speller applications

The primary concerns in regards to the frequency used in the stimuli are; the refresh rate used to display the frequencies and their associated harmonics, whilst being able to avoid overlapping between them – as an example, choosing 10Hz as a primary frequency would have a harmonic of 20Hz, making 20Hz a poor choice for another primary frequency. In two different works, [40, 24], the authors propose two different ways of improving the discrimination between classes for a SSVEP-based speller using a keyboard stimuli. The authors in [40] propose a method called Joint Frequency-Phase Modulation (JFPM) where different phases can be added to the frequencies to compensate for the small difference gap between them. To avoid frequencies overlap with harmonics, [24] use 28 spaced frequencies (with a range between 6.10 and 11.28Hz) where none of the frequency overlap with other harmonic frequencies. Instead of having alphanumerical symbols for stimuli speller, the stimuli in [166, 235] displays 10 numbers and 2 symbols as a virtual phone keypad.

Even though the work in [40] shows the ability of a BCI speller to use up to 40 frequency classes, the process of collecting data for calibration can be tedious as each class needs a number of samples to train a classifier. Therefore, multi-layer spellers have been used to keep the frequency classes to a small number, whilst able to have a more significant number of stimulus for different characters [238, 172, 203].

SSVEP response has the features where the recorded cortical signals contain the peak frequency and its harmonics. Certain applications allow for the same frequency to be reused for another character at a different time. Multi-layer spellers work by grouping some characters into separate groups, where the first layer of speller allows choice of the group that contains the intended symbol. The next layer separates the character in the chosen group to either a smaller group or a single character. This process is repeated until a final symbol is chosen.

The authors in [238] apply the JFPM method on six target frequencies where each frequency is not designed for only a character. During the offline data collection stage, all same-sized boxes were labelled as number one to six. For the spelling tasks, two layers of speller interface were used, where in the first layer, the six boxes consist of five characters per box. The characters in the selected box are separated into individual boxes in the next layer where one can be chosen by subject as a target character.

Similar work is presented in [172] where the spelling process consists of three layers, where each layer has five target frequencies. This work provides 58 characters in the stimuli; hence the more complex layer as compared to previous work. Three layers speller have been adopted by [203] for a web browser-driven application. The web speller can be configured by a user whereby they can manage and choose four out of seven frequencies (6.0, 7.5, 8.0, 8.5, 9.0, 9.5, and 10Hz), as well as the word to be spelt. Nine characters are grouped in the first layer for three boxes, and the fourth box is for the 'DELETE' symbol. Once the user manages to choose the correct target, the second layer of the interface appears with three boxes consisting of three characters each (from the targeted box in the first layer) and the 'BACK' symbol in the fourth box. Finally, the last layer consists of one character per box for 3 boxes and the 'BACK' symbol in the fourth box.

SSVEP-driven Robot

As SSVEP stimuli can be used to display an icon or specific symbol, it is useful for BCI-based robotic applications, where the different frequencies can be interpreted as different commands.

Commonly, SSVEP stimuli consist of a number of flashing boxes, which are used to navigate a robot to move around in an environment. The stimuli can be displayed either on a monitor, a tablet or even a physical LED with different sets of frequencies. However, as discussed previously, there is no standard way of designing SSVEP stimuli in term of frequency selection, the size of the boxes and the stimuli display [151, 3].

Application	**Frequency (Hz)**	**Control**
Mobile robot in an open office [58]	{37, 38, 39, 40}	forward, turn right, turn left, stop
Humanoid robot for obstacle avoidance [96]	30 with stimulation phases [0°, 300°] with step of 60°	forward, backward, shift right, shift left, turn right and turn left
Mobile robot in real environment [64]	{5, 6.666, 7.5}	forward, turn right, turn left
Mobile robot for obstacle avoidance [137]	{8, 10, 12}	forward, turn right, turn left
Meal assistance using robot arm [185]	6, 7, 8	bowls position 50, 60, 65cm
Pick and place robot arm [42]	{30, 31, 32, 33}	three shapes and undo
Pick and place robot arm [41]	[8, 15.2] with step of 0.3Hz	position of 25 targets
Two-layer mobile robot navigation [218]	[6.0, 20.0]	three commands and mode switch in both layers
Navigation of humanoid robot [252]	6.25Hz, [11, 19] with step of 1Hz	two-layer stimuli using captured regions
Controlling humanoid robot behaviour [260]	{5, 6, 8, 10}	three-layer stimuli
Controlling humanoid robot behaviour [257]	five frequencies from {4.615, 6.667, 8.571, 12, 15, 20}	two-layer stimuli

Table 2.2: SSVEP-based application for robotic applications

As summarised in Table 2.2, there are two main types of robot application via SSVEP: basic control and multi-layer control. Basic control stimuli consists of three to four different frequency boxes; each indicating a basic control such as forward, turn left, turn right, or stop movement. This system requires the subject to execute the command over time to navigate the robot as there is no predetermined path.

Physical Light-emitting Diodes (LED) can be used to display high-frequency stimuli for a less fatiguing experience to subjects. The work in [58] presents a robot application using four

LED attached to a computer screen. Subjects are able to monitor the environment of the robot in real-time displayed on the screen. Instead of using LED for high-frequency stimuli, the work in [96] uses 30Hz frequency signals on six stimulus boxes. They manipulate the stimulation phases by displaying different dark-bright sequences for each stimulus. Each target stimulus is used for different navigation commands to allow a humanoid robot to avoid obstacles in an office environment.

Similar applications are presented in [64, 137] for mobile robot navigation. However here, the stimuli are displayed on an LCD monitor screen. The robot in [64] is equipped with a wireless camera to provide subjects with live video feedback. Subjects in [137] can view the robot condition themselves and are required to navigate the mobile robot from a starting point to destination point whilst avoiding obstacles.

The studies using SSVEP to manipulate a robot arm in [185, 42], could be considered as a basic control application as only three commands are used. However, the approach in [185], is interesting as instead of selecting one of the typical basic movement commands (left or right), a predetermined complete path is chosen to allow for meal selection in a meal assistance application. In [42] pre-trained object recognition is used to identify three different objects and 25 locations. Subjects are required to select an object at a time to move the robot arm to pick up the selected object and place it at the desired location. Taking advantage of better technology, this paper uses a screen monitor with 120Hz refresh rate to display high-frequency SSVEP-stimuli. Similar work in [41] shows a control robot arm for a pick and place task for 25 locations where each frequency is corresponding to a location.

Multi-layer stimuli are useful for robot applications, as different robot systems have complex control modes. In [218], the authors use a two-layer stimuli application for two different modes: drive mode and camera mode – with four commands for each mode to navigate a mobile robotic car (MRC). The application has a live video feedback where subjects can control the pan-tilt-position in camera mode. The MRC comes with a sensor that stops automatically at a turning point and makes small corrections when it detects the line. They also developed a calibration software coined ‘SSVEP Wizard’ to determine the key parameters for the SSVEP stimuli for each subject. This software chooses the best four frequencies for each subject to be used as the SSVEP frequency stimulation.

The image frame captured from a robot camera can be used as stimuli as well, [252] where the frame is divided into nine equal regions, where each region is flashing at a different frequency.

After selecting a region in first-layer stimuli, the selected region is further divided into nine equal regions to enable more precision selection. The specific region is used to navigate the robot to complete a trajectory.

The authors in [260] design a multi-layer SSVEP stimuli to control a humanoid robot. The first-layer stimuli is associated with the robot state, either sitting, walking or running into an obstacle, which is decided by the robot controller. Under the sitting state, subjects can select to observe surrounding, monitor the robot hardware state, or command the robot to stand up. If the robot encounters an obstacle at any time, subjects can choose to either avoid the obstacle or to interact with it (moving it for example). During normal walking state, there are two main options for subjects which form the second layer: exploration or interaction – each having four options to be performed.

Similarly, the authors in [257] propose to control a humanoid robot using behaviour-based SSVEP to walk in a cluttered environment, and ultimately approach and pick up a target object (Green balloon). The stimuli consist of five frequencies, four of which are for a two-layer stimuli for top-menu and sub-menu movements, and one frequency dedicated for returning to the previous menu. The frequency set is subject-specific, where during the calibration stage, the five most responsive frequencies are selected for every subject. By selecting any one of the top-menu elements, the subjects can control another four behaviours of the robot in the sub-menu. The user interface of the system consists of five stimuli, a display for live video feedback in the middle and a display for the current posture of the robot. A camera embedded on the head of the robot is used as location feedback, where the subject can monitor the location and the surroundings of the robot.

Other Applications

Aside from the applications mentioned above, there are additional uses of SSVEP-based BCI in the literature. These include,ICU communication [56], virtual reality maze game [240], remote control vacuum cleaner in a virtual environment [217] and user authentication [248].

Research for an assistive application for Intensive Care Unit (ICU) patients has been performed in [56], where SSVEP-based stimuli were used for communication purposes in an ICU environment. To make it comfortable and easy to be used by patients, the system used a portable

wireless wet-EEG headset for streaming the cortical signals and a tablet to display the stimuli. The stimuli consisted of a default screen with an icon to call a nurse. When a subject fixated on the icon, the screen displayed the main menu comprising four boxes, one at each corner of the screen, flickering at different frequencies (8.5, 10.0, 12.0, 15.0 Hz). Each box contained a sub-menu where there are two or three icons to be chosen for different requests and a home button icon to go back to the main menu. This design helps to eliminate the number of frequencies limitation by having different icons appear at a different time and frequency reuse.

Instead of using SSVEP-based stimuli to control an external physical device, they can be used in virtual life, either for a game or for research purposes. A mobile robot was controlled virtually via an SSVEP-based robot GUI [240], using frequencies of 7 Hz, 8 Hz, 9Hz and 10 Hz. Subjects were required to navigate a mobile robot to move forward, backward, turn clockwise and counter-clockwise to complete a maze. Additionally, a virtual vacuum cleaner was used in [217] as a comparison study on SSVEP-based stimuli for two types of displays, on a head-mounted display (HMD) with 90Hz refresh rate and on a 120Hz refresh rate laptop screen. Subjects were required to navigate the robot to collect 10 dust piles by looking at one of the frequency stimulus boxes at a time for both studies. The frequencies used were 7.5Hz, 10Hz and 6.0Hz for turn left, forward and turned right respectively. A similar study on HMD is presented in [25], where frequency-based stimuli were displayed on HMD for subjects to walk in a virtual environment.

As it is well known that EEG signals contain highly unique features for every subject, [248] investigate the consistency of the individual subject pattern for authentication purposes across eight subjects. By using a 40 character alphanumeric stimuli, with a frequency range of [8.0, 15.8] with a step of 0.2Hz and phase between [0.0 π, 1.5π] with a step of 0.5π, EEG signals are collected on two different days. This is to ensure that subjects' identity can be recognised accurately across different days.

As demonstrated by the previous works, SSVEP-based response is a versatile way to enable BCI applications, where it is possible to assign different commands to a frequency providing that the frequency does not appear more than once on different stimulus at a same time. SSVEP-based stimuli have been shown to be a pragmatic approach used in a wide variety of tasks, such as alphanumeric symbols, icons, arrows and even images as long as the frame is flashing at a particular frequency. However SSVEP has limitations, such as the number of available frequencies is limited by monitor refresh rate and the chosen frequencies overlapping with harmonics of other ones [177, 247].

Despite this and inspired by the ability of SSVEP stimuli, in this thesis, we explore the potential of improving the stimuli display for a real-time natural environment application. The details of the SSVEP stimuli are further discussed in Chapter 3 and Chapter 4 of this thesis.

2.3 Bio-Signals Pattern Recognition

To establish reliable communication in BCI application, the translation of raw EEG cortical signals to a specific desired decision is a crucial task [177, 160, 200]. The cortical signals measured via EEG is a non-invasive process, meaning that it is not directly measured from the cortex but through the skull. As the skull is not a good electrical conductor, the electrical activity measured suffers from poor spatial resolution with low SNR [143, 201]. Not only that, these bio-signals come with unwanted signals and artefacts such as eyes blinking, muscles movement from head and neck, inadequate electrode allocations and power line noises [230, 216]. Therefore it is a challenging task to process the signals in a proper manner where the correct pieces of information are extracted and can be used to decode the signals accurately [136].

Subsequently, signal processing is one of the primary components in the field of BCI as it acts as the translation layer between the raw EEG cortical signals and a specific desirable decision or application [160]. The accurate representation of EEG translation can be achieved by assigning the raw bio-signals to targeted classes by learning relevant patterns in the signals [177, 201]. The process of learning the patterns can be done autonomously in the field of Machine Learning (ML) where it investigates the use of machines to learn patterns in datasets automatically via the use of specialised algorithms [208].

ML is a data-driven process which relies upon the input dataset for learning the correct result for a given task. This learning process allows the family of algorithms to perform a variety of tasks on the dataset including classification (the process of assigning an unknown data example to a discrete label) and regression (the process of predicting a continuous range output variable for a given data example based on previous trends within the data). The datasets used for learning are often partitioned into at least two groups: training data to be used as input to the training algorithm and the testing data to be classified or predicted based on the trained model [81].

As broadly defined, the field of machine learning can be divided into two main categories: supervised learning and unsupervised learning. In supervised learning, the data is provided with a specific target output, also known as a label. A supervised ML algorithm exploits these labels to learn the best solution for the given task. In contrast, unsupervised learning algorithms require no such labels or target outputs for the input dataset. Instead, unsupervised algorithms automatically learn to identify patterns and trends within the data [21].

In this section, we discuss the different approaches presented in the literature to improve the classification task on bio-signals, especially on SSVEP-based applications. We separate the traditional ML approaches and Deep Neural Network (DNN) approach.

2.3.1 Traditional Machine Learning Approaches

Commonly, in a BCI signal processing pipeline, there are three primary stages; pre-processing, feature extracting and signal classification (Figure 2.2) [216, 177, 4]. The pre-processing stage is performed to eliminate irrelevant noise from the raw bio-signals to focus on the desirable range to ensure they are useful and distinguishable [230]. There are different types of pre-processing depending upon the electrode locations, neuro-physiological response used for collecting the bio-signals and the type of features to be used to represent the signals [136, 142, 201].

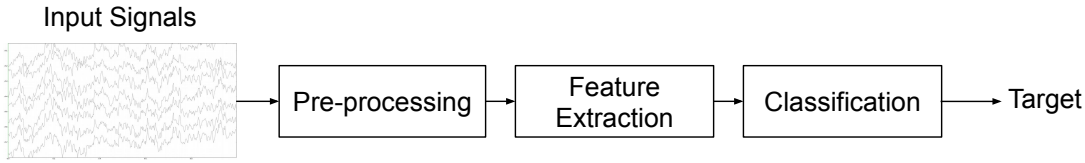


Figure 2.2: Traditional BCI signal processing pipeline

Typically, a bandpass filter is used as a temporal filter to eliminate other electrical activity from different parts of the body (such as eye blinking, muscle movement) and focus only at the targeted frequency range [225, 33, 177, 146, 36, 160, 63, 206]. If the range is significant and includes power line noise, a notch filter can be used to remove a specific unwanted frequency [33, 177, 146]. A spatial filter approach can be used as well to enhance the signals from a specific cortical region, for example using Common Average Reference (CAR) to reduce the common noise across electrode channels [225, 33, 36, 206] or using Common Spatial Pattern (CSP) [136, 177, 101].

The filtered signals are then utilised to extract relevant and distinguishing features from the signals [142]. Extracting and selecting features in the signal processing domain is a crucial step as signals with similar features are classified together [177]; hence the selected features should be distinguishable between different classes [143]. Usually, the traditional ML features are manually hand-crafted as they can be represented in various ways using different extraction methods depending on the paradigm and the task-relevant information [143].

The most common feature types extracted from bio-signals are spectral and temporal features [142]. Spectral features are representing the signals by extracting the power band of a given frequency band of the signals [33, 177, 142], which can be done by using a Power Spectral Density (PSD) approach [4, 201]. Temporal features can be useful for classifying time variation tasks (such as P300-based ones) and non-stationary characteristics of EEG signals. Hence instead of focusing on spectral features, approaches such as Short Time Fourier Transform (STFT) [230, 177] and Wavelet Analysis [169, 101], are able to generate frequency representation based on each point in time of the bio-signals [33, 5]. Often a spatial filter is used as feature extraction as well to extract spatial features; for example, CSP is used to enhance the signals by maximising the representation of the signals of one class to another [3, 142, 146].

After this feature extract stage has been performed to create a representation of the cortical signals, a mapping between it and a set of class labels or catagories is learned using a classification technique [216]. Some of the conventional ML-based techniques used to classify SSVEP-based signals are Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Riemannian-based and CCA, to name a few [225, 40, 177, 142, 3].

One of the most frequently used classifiers within the BCI field is LDA [33, 33, 177, 142, 206], a simple linear classifier created by computing the separating line with the highest discriminatory power between classes by reducing the dimensionality, creating a low computational cost [33, 29, 216, 177, 65]. Although LDA can easily classify data which is inherently linear, the classification of non-linear data is non-accurate [29].

SVM is another linear classification method aiming to separate the classes of the neural signals upon finding the maximum margin hyperplane or linear separator between classes [212, 216, 5, 36, 88, 243, 247]. SVM works well for a two class, otherwise known as binary, classification task [136, 212], relying heavily upon the quality of the input features [33] and demonstrating a high computational complexity [101]. SVM works on non-linearly separable problems as well

when kernel functions are used and can support multi-class classification using strategies such as one-against-all and one-against-one [95].

As a consequence of SSVEP-based signals being identified via the associated fixation frequency, many works employ CCA-based techniques to perform the frequency classification. CCA is a multivariate statistical approach that finds which one of stimulation frequency used has the maximum correlation with the cortical signals [254, 40, 116, 232, 36, 210, 236, 168, 76]. Since it uses stimulation frequencies as the reference dataset, for most of the time it is used, this approach does not require any calibration stage, making it appealing for SSVEP-based applications [118].

However, as the approach totally relies on the frequency and its harmonics, it is a less robust technique [36] to the point, some offline calibration still needed to increase performance and reduce the intricate procedure [32]. The chosen frequencies need to be carefully selected as CCA does not perform well if the frequencies or associated harmonics overlap with each other [150, 102, 43]. Besides, the classification performance depends on the assumption that all the bio-signals have similar characteristics, with high PSD values leaving the BCI session to be a very rigid protocol, which is not suitable for robust daily application [65, 43].

In order to alleviate limitation due to the feature selection process, Riemannian-based classification approaches have received attention in regards to the elimination of feature selection stage [20, 245, 141, 102, 43]. Riemannian geometry distance is used as features, where the distance is determined by mapping the covariance matrix of bio-signals on a tangent space of a curved geometrical space [19, 20, 43]. The distance can be used to classify the bio-signals using either traditional classifiers such as SVM [20] or LDA[19], or simple a classifier like Minimum Distance to Mean (MDM). MDM classifies a signal by computing the minimum distance of the geometric mean determined by the feature distance [19, 102, 43]. However, due to computation of the square of the geometric distance, the approach may face numerical instability and it can increase the computational complexity [142].

2.3.2 Deep Neural Network Approach

The main attributes of traditional ML algorithms depend upon proper feature extraction and selection to be used as input. However, these vitally important stages of getting the optimal

results are done manually based on human engineering by selecting the combination of algorithms for the application and paradigm [247].

The development of signal processing for BCI is essential to ensure that all the required features are included, but at the same time to avoid redundancy [216] that might harm the classification stages. The discriminative power can be improved by having autonomous methods to extract and select the proper features, as relevant features in EEG data is not always a priori information [208].

Recently representation learning has been explored as a way to combine both the feature extract and learning processes. Representation learning can take raw data or pre-processed data – depending on task, as input and automatically discover the highest quality representation [121]. Deep Neural Network (DNN) is a sub-field of machine learning, based on representation learning, composed of non-linear but straightforward units with multiple layers of processing to learn the representation of raw datasets [81].

While training for a particular task, a DNN model computes the output score for a given loss or objective function and measures the error against the ground truth. Based on the error, the parameters (weights) are adjusted to minimise this error. The proper adjustment of the weights is determined by computing the gradient vector with respect to the loss. The process is repeated with numerous passes through the data to get the desired result [121].

Deep Neural Networks are known machine learning method in various areas such as natural language processing, speech recognition and advertising to name a few. However, perhaps the most widespread application being its use in the classification of images [208, 121]. Nevertheless recently, the approach has been used to improve the classification of various aspects of EEG data [115, 226]. The ability to classify EEG signals end-to-end by removing the requirement for hand-crafted feature extraction provides a critical advantage as the potential exists for salient EEG signals or features to be excluded or missed when using traditional machine learning-based approaches [120, 235].

Convolutional Neural Network

Convolutional Neural Networks (CNN) have emerged as the most widely used DNN models in classifying EEG data. CNN are a feed-forward neural network that learns to differentiate between

classes in data by learning unique features across multiple layers of convolutional transformation [121, 142]. In the convolution layer, the input is convolved via kernels (filters) to obtain feature maps from low-level features in the first layer to high-level features in deeper layers [81, 175, 208]. This process removes the requirement for hand-crafted feature extraction as well as common signal pre-processing steps, as raw data samples can be used as direct input to the model [81, 208]. This property facilitates the model to learn complicated mappings within the data by including all the potential distinguishable features of EEG signals between classes that are possibly overlooked by traditional approaches [120].

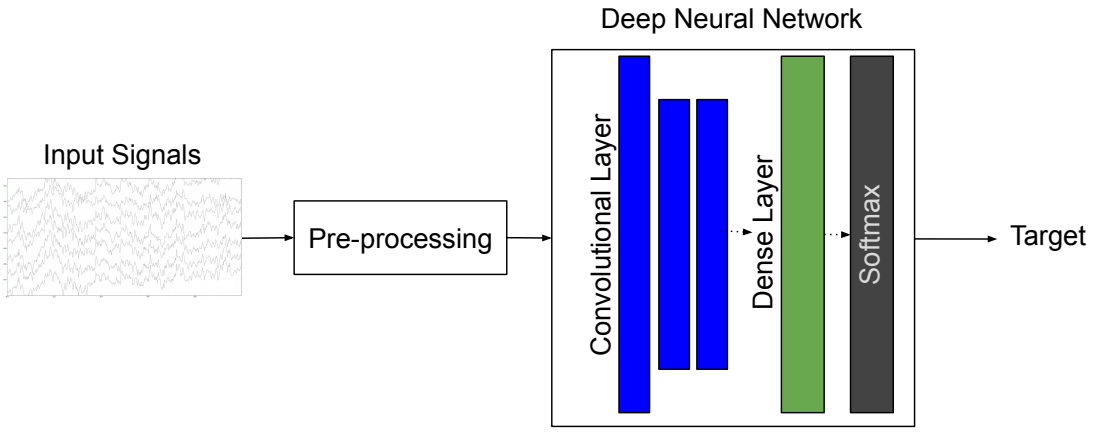


Figure 2.3: Example of Deep Neural Network pipeline on BCI signal processing via Convolutional Neural Network

There are a growing number of different BCI applications that have adapted CNN to classify EEG-based bio-signals [201]. As an example, CNN have been used in [175] to decode the ambiguous cortical signals such as EEG/LFP (local field potential) signals from self-paced left and right-hand squeeze tasks. Music perception and imagination can be decoded using CNN [220], where the cortical signals recorded can be differentiated whilst the participant is listening to certain music stimuli. In [34], image patterns from reaching a target task via BCI-VR (virtual reality) can be learned via the use of a CNN. The images are constructed by concatenating subfrequency components of cortical signals measured via the use of EEG.

Another work using CNN to classify an EEG image extracted from a MI task via short-time Fourier transform (STFT), is presented in [200]. They compared CNN performance in classifying left and right-hand imagination movement based on two different optimisers; Adam

and SGD (Stochastic Gradient Descent). A traditional ML, Filter Bank Common Spatial Pattern (FBCSP) was used as a baseline to be compared against different configurations of of CNN when classifying four classes MI tasks. These configurations are Deep CNN (four convolutional-max-pooling layers), Shallow CNN (one convolutional-max-pooling), Hybrid CNN and Residual CNN (ResNet) [208].

A combination of CNN and Stacked Auto-Encoder(SAE) has been used in [223] for EEG MI classification on the left and right-hand imagination movement. In the presented CNN-SAE approach, the output of the convolutional layer is the input of the SAE network. A CNN model, EEGNet has been introduced by [120] for EEG classification across paradigms. The paper includes four datasets for four different paradigms (P300 Event-Related Potential, Error-Related Negativity, Movement-Related Cortical Potential and Sensori Motor Rhythm) from different sources with different data sizes.

CNN based models have also been utilised to classify SSVEP-based EEG signals, for example for a virtual reality application in [25]. A comparison study on DNN approaches in classifying five-class SSVEP signals against traditional ML approaches has been performed in [226], where classifiers include k-Nearest Neighbour (k-NN), decision tree, Multi-layer Perceptron (MLP), SVM, AdaBoost, CNN and Recurrent Neural Network (RNN) with Long-short Term Memory (LSTM). Importantly, the author suggested that the DNN ability in classifying EEG signals improves as dataset size is increased.

Owing to the EEG data availability and access problems, a compact-CNN has been proposed by [235] to classify ten frequency SSVEP classes. Compact-CNN is a 2D convolutional network consisting of temporal convolutionals, depthwise spatial convolutional and separable convolutional. This proposed approach not only ensures the efficient extraction of frequency-specific features but also combines information across temporal and spatial filters. The proposed DNN method is compared against CCA and extension version of CCA; combined-CCA and is shown to outperform them.

A non-stationary SSVEP application, such as controlling an exoskeleton online via a CNN, has been proposed in [115]. Five-class SSVEP frequencies were decoded and compared between CCA, Multivariate Synchronization Index (MSI) and CCA with k-Nearest Neighbours (CCA-KNN). These were used to compare the classification result with three proposed neural-based methods: CNN-1 (3 layers network), CNN-2 (4 layers network) and fully-connected MLP. For

CNN-1, the authors proposed two hidden layers (C1 and C2); both are composed of eight maps with 120 for C1 and 110 for C2 and an output layer that has five units (5 frequency classes). This layer is fully connected to C2. The authors introduced the third hidden layer (F3) in CNN-2 in addition to the layers in CNN-1. F3 consists of three units and fully connected, which are used for the learning visualisation on CNN-2 properties. The standard MLP consisted of four fully connected layers: an input layer, two hidden layers and an output layer.

In another work, frequency-based features are used as input for two convolutional-layer CNN to classify a five frequency multi-layer SSVEP speller [172]. There are two types of features computed using FFT on the various number of points and time windows, where the first type is computed by normalising FFT features between zero and one and the second type is by dividing the FFT features with maximum FFT power points. CNN performance in the detection is compared against CCA and Least Absolute Shrinkage and Selection Operator (LASSO), where CNN outperformed these two approaches.

Limitations of Deep Models

Perhaps the major drawback of deep learning is that it represents a 'black box' approach, where there is currently no fundamental theory as to why it so successful. Due to this, it can take time to design a deep network that works well for a given task. One must frequently adopt a trial and error approach for choosing key parameters for the network [81].

In addition to this, to achieve the best performance using deep learning techniques, large datasets are required for training. As deep learning is currently mostly utilised in supervised learning, the large datasets must additionally be labelled. Due to the requirement of a massive dataset as input, deep learning is an expensive approach to train from a computational vantage point especially in using bio-signals. Collecting high quality EEG data has proven difficult as the experiment procedure is time-consuming and expensive [53, 253, 145].

Additionally, the complexity model also increases the demand upon the computing resource. Due to such demands, Graphics Processing Units (GPUs) are commonly utilised for model training, although the training time for large models can still be measured in days [81]. Even though the field of deep learning has proliferated recently, the research on unsupervised deep learning has received less attention. It is an essential area for research to be performed on to fully utilise the unlabelled data for unsupervised learning tasks.

Throughout this thesis, we focus on the use of DNN, specifically CNN to perform the classification of EEG bio-signals. In Chapter 3 of this thesis, we compare different machine learning techniques and the details of our proposed CNN.

2.3.3 Bio-Signals Pattern Generation

In machine learning, generative models have long been used to generate entirely new and realistic data points which match the distribution of a given target dataset [205]. Generative models have been proven very powerful within the context of unsupervised learning where the model learns a hidden structure of the data from its distribution to generate new data samples within the same distribution [99]. This generated dataset often contains enough variation to support the down-stream training of a secondary model [145].

A generative model can not only can be used for pattern recognition but it is can also be used to generate synthetic data that has real data characteristics. This could potentially be used as a form of data augmentation to help improve the performance of any down-stream data classification task [72]. Recent work on neural-based models such as Generative Adversarial Networks (GAN) and Variational Auto-Encoders (VAE) have demonstrated that these models are highly capable at capturing key elements from a diverse range of datasets to generate realistic samples [82].

Neural-Based Generative Models

Auto-encoders have long been used as a method of creating a low-dimensional representation z of data using an encoder model, which can be used to reconstruct the original data with minimal errors via a decoder model [92]. However, traditionally they cannot be explicitly used to generate new data samples based on the learned data distribution. Variational Auto-encoders (VAE) utilise ideas from Bayesian inference to produce a more expressive data representation, whilst also having the ability to generate new data samples [108, 198]. Unlike non-probabilistic auto-encoders [92], a VAE does not learn a fixed value for each element in z but instead each element is sampled from a probability distribution before being passed to the decoder model. This has been shown to produce a more semantically meaningful representation, where individual dimensions in the hidden space can correspond to tangible elements in the dataset, such as facial

expression in a dataset of human faces [119]. As the decoder model of the VAE is trained to take a sample from a Gaussian distribution and produce a realistic output, it can be used to produced new data by simply sampling points in the z distribution and reconstructing them via the decoder.

In contrast to Auto-encoders, Generative Adversarial Networks (GAN) [82] are capable of producing semantically sound artificial samples by inducing a competition between a generator (G), which attempts to capture the distribution, and a discriminator (D), which assesses the generator output and penalizes unrealistic samples. Both networks are trained simultaneously to achieve an equilibrium. However, training a GAN is known to be challenging with pervasive instability issues [8]. One such issue stems from the discriminator rapidly reaching optimality and effortlessly distinguishing between the fake samples output by the generator and samples from the real distribution. This will lead to a lack of meaningful gradients for training, effectively ceasing any progress towards the equilibrium.

The Wasserstein GAN (WGAN) is consequently proposed in [9] to rectify some of the issues associated with training a GAN. The Wasserstein-1 metric is used to measure the distance between the real and model distributions. Also known as the Earth Mover's distance, ($EM(p, q)$), this metric is the minimum cost of moving distribution elements (earth mass) to transform a distribution q to distribution p (cost = mass $\times$ transport distance).

Within the existing body of work in the literature, there are instances of generative models being used to create synthetic EEG data. For instance in [53], the authors propose deep EEG super-resolution using a GAN. The model is applied to a small number of EEG channels with the goal of interpolating other channels using motor imagery dataset from [159]. The super-resolution data (SR) is generated via a WGAN with convolutional filters using the low-resolution data (LR) from the dataset by down-sampling the EEG channels by scale factors of two and four for two different experiments. The channels removed from the down-sampling processes are used as high-resolution data (HR) to compete against SR in the discriminator. They evaluate the performance of the SR by performing classification and comparing the accuracy with the classification performed using HR. The authors conclude that SR is capable of producing high spatial resolution EEG signals from low resolution signals.

In [253], synthetic EEG signals are generated by inverting the artificial Time-Frequency Representation (TFR) obtained from a conditional Deep Convolutional GAN (cDCGAN). The

Wavelet Transform is used to obtain the TFR of the signals which are to be used by the cDCGAN to generate the artificial TFR of the EEG signals. The BCI competition II dataset III [209] was used as the training EEG data. They evaluate the efficacy of the synthetic data by comparing the classification accuracy of the model trained on real data, synthetic data and a mixture of real and synthetic data with different ratios. Importantly, using additional synthetic data, the accuracy of the model improved by up to 3%.

In [145] and [191], a GAN was used to generate synthetic emotion recognition EEG data. The authors in [145] generate EEG in the form of differential entropy from a noise distribution using a conditional WGAN with two emotion recognition datasets. They evaluate the performance of the generated data by combining the synthetic and real-world data to train a classification model and compare against a model solely trained on real-world data. The addition of the synthetic data leads to an improvement in the accuracy of up to 20% for different datasets.

In [89], the authors propose using the improved WGAN with Gradient Penalty (GP) for a single channel of the EEG data for motor imagery tasks. The generator consists of one linear layer, six convolutional layers, where each layer consists of an up-sampling operation, two convolutions and one fully-connected layer. The authors evaluate the performance of four models, three different up-sampling methods (nearest-neighbour, linear interpolation and cubic interpolation with convolutional down-sampling) and down-sampling via average pooling with the original WGAN-GP. All the models are then assessed using four evaluation metrics; inception score, Frechet inception distance, Euclidean distance and sliced Wasserstein distance. All four models are demonstrated to outperform WGAN-GP.

In Chapter 5 of this thesis, we explore the potential of the generation of new synthetic EEG data using a selection of customised neural-based generative models to boost classification accuracy on real datasets. We specifically focus on dry-EEG data containing Steady State Visual Evoked Potential (SSVEP) signals.

2.4 Limitations of BCI

This chapter has presented a literature review on a variety of crucial components of BCI. In this section we highlight and summarise some limitations concerning the efficient applicability and usability of this technology:

- *Rigid EEG-based Experimentation* - Commonly, the EEG protocol via conventional wet-EEG technology can be rigid, consume much time and typically only suitable for use in a laboratory setting (Section 2.1.1). This nature of experimentation makes EEG BCI experiments expensive and challenging to operate within the usually short amount of time experimental subjects can perform the experiments [53].

- *Frequency Limitation* - As SSVEP is frequency-based stimuli, selecting the set of frequencies is crucial for any SSVEP-based application. Using low frequencies as stimuli is a straightforward implementation concerning the monitor refresh rate. However it can lead to fatigue and might even cause epilepsy to particular subject [103]. It is important to note the gap between the frequencies used, as any overlapping with other frequency harmonics will affect the quality of signals [177, 247]. Additionally, the design of stimuli can be rigid in terms of the size and pattern of the bounding box used, placing further constraints on flexibility in expanding usability (Section 2.2.2).

- *Hand-crafted-based classifier* - As measuring cortical signals through scalp is to susceptible noisy, many traditional approaches focus on improving the quality of the signals. This can lead to a very rigid protocol and specific solutions being developed to accommodate the hand-crafted features (Section 2.3.1). This increase the complexity of a given BCI application, for example, mandating the use of a system to personalise frequencies or the electrode sensors used [236, 182].

- *Small dataset* - Collecting high quality EEG data has proven difficult as the BCI experiment procedure is time-consuming and expensive [253, 145, 53].

- *Calibration Stage* - It is commonly understood that EEG signals contain unique patterns for specific subjects [141, 2, 56, 117, 248]. This can result in an approach being highly subject and session variant, leading to a long calibration process and further restricting its real-world practical applications.

2.5 Relevance to Contributions

In this section, based on the current literature and limitations related to BCI application, we outline the novel contributions of this thesis made in the following chapters.

Despite being the most prominent data acquisition device used in the BCI world, EEG-based experimentation is known for a rigid and lengthy procedure to ensure the quality of the bio-signals gathered, thus leads to low applicable and usability in a real-world environment [160]. The recent development of portable dry-EEG technology is covered in Section 2.1.2 where such technology eliminates such a rigid procedure and reduces the preparation time for a BCI experiment, although it raises issues including the stability and accuracy of the signals. Employing this, we detail the SSVEP data collection via dry-EEG technology in Chapter 3, exploring the potential of having a system that can be used outside of the laboratory. Taking advantage of the ability of DNN to improve the accuracy of the classification task, as discussed in Section 2.3.2, we explicitly consider an end-to-end approach using a novel CNN model, to tackle SSVEP-based dry-EEG signal classification challenges.

As discussed in Section 2.2.2, SSVEP stimuli are versatile as they can be used in many applications, from alphanumeric symbols, icons, arrows and even images as long as the frame is flashing at a certain frequency rate. Inspired by the ability of SSVEP stimuli as a medium for a BCI system, in this thesis, we explore the potential of improving the stimuli display for a real-time natural environment application. In contrast to earlier work, in Chapter 4 we design a novel stimuli by taking advantage of the on-board camera on our humanoid robot and employ the use of a high-performance scene object detection model for variable BCI stimuli, embedded within the scene video feed in a real-time application. As there is a limitation on the number of SSVEP frequencies that can be used (Section 2.4), we assign different commands to a frequency class as long as the frequency does not appear more than once at a time.

The ability of DNN to autonomously learn unique features from bio-signals is beneficial in a classification task and boosts the performance on dry-EEG signals. As mentioned in Section 2.4, it is essential to have access to large quantities of data for training such methods. However the difficulty of collecting high-quality EEG data is that it can be expensive and difficult for various reasons, severely hindering the applicability of the system. In Chapter 5, we explore the potential of using various unsupervised DNN approaches to generate synthetic EEG signals which capture realistic elements from real-world bio-signals. The generated synthetic data can then be used as a form of data augmentation to help improve the performance of any down-stream data classification task.

Additionally, EEG data is known to be highly subject and session variant, which leads to severe calibration requirements for any EEG-based BCI application at the beginning of any task

to measure a specific subject and session conditions (Section 2.4). With the high capability of DNN in capturing bio-signals features, the issue of subject bias should be taken into consideration while dealing with such data in a learning-based approach. In Chapter 6, we include a novel investigation on generating highly subject-invariant synthetic EEG signals via generative model subsequently used to classify unseen subjects.

Overall, in this thesis, with recent technology and the improvement of Deep Learning, we explore the potential of EEG-based BCI applications - with a focus on using SSVEP to enable a robust system that can be used in daily-life without compromising the performance of the system.

Chapter 3

Classifying Dry-EEG Bio-signals

Dry-EEG is one of the state-of-the-art data acquisition methods used for BCI applications. It is an improved alternative to traditional wet-EEG which rely on the inconvenient application of conductive gel to ensure that a high-quality signal can be obtained. This advanced technology improves the experiment protocol experience and reduces the time taken, thus alleviating the unpleasant process for the experimental participants. While improving the usability of EEG within the BCI context, dry-EEG suffers from inherently reduced signal quality due to the lack of conduit gel, making the classification of such signals a significantly more challenging task.

In this thesis, we elevate dry-EEG technology as the data acquisition method for BCI applications to prove the ability of such technology to stream and record human bio-signals. This chapter focuses on integrating a dry-EEG headset with simple basic SSVEP stimuli to demonstrate the competency of the headset to obtain appropriate information. The utilisation of SSVEP (introduced in Section 2.2) allows people with severe physical disabilities such as Complete Locked-In Syndrome or Amyotrophic Lateral Sclerosis to be aided via a BCI application, as it requires only the subject to visually fixate upon the sensory stimuli of interest.

In this chapter, we propose a novel Convolutional Neural Network (CNN) approach for the classification of raw dry-EEG signals without the need of any prior data pre-processing. We incorporate different machine learning techniques to evaluate the performance of the classification task against our proposed methods. We also detail the experimentation protocol, which includes

the procedure of using a dry-EEG headset, the stimuli design, the proposed network architecture and the integration setup between all components. It should be noted that in later chapters, we use these protocols again, with modifications to some of components.

Some portions of the work presented in this chapter have been published in the following peer-reviewed publication:

> *Nik Khadijah Nik Aznan, Stephen Bonner, Jason Connolly, Noura Al Moubayed, and Toby P Breckon. On the Classification of SSVEP-based dry-EEG Signals via Convolutional Neural Networks. In* International Conference on Systems, Man, and Cybernetics*, pages 3726–3731. IEEE, 2018*

3.1 Introduction

In this chapter, we investigate the use of a deep neural network, specifically a CNN, to perform the classification of SSVEP frequencies in dry-EEG data. CNN are a subset of neural networks, which learn to differentiate between classes in data by extracting unique features across multiple layers of convolutional transformation [121]. In the convolution layer, the input is convolved via kernels (filters) to obtain feature maps [81]. This process removes the requirement for hand-crafted feature extraction as well as common signal pre-processing steps, as raw data samples can be used as a direct input to the model [81, 208]. This property provides a critical advantage as the potential exists for salient EEG signals or features to be excluded or missed when using traditional pre-processing based approaches [120].

We evaluate the performance of our proposed CNN architecture at classifying dry-EEG SSVEP signals across a four-class stimuli problem collected from a single subject and highlight the vastly superior performance when compared to baseline classifiers including the Support Vector Machine (SVM) [20], Linear Discriminant Analysis (LDA) [19], Minimum Distance to Mean (MDM) [19] and a Recurrent Neural Network (RNN) [27]. Furthermore, we explore the use of the same CNN architecture to examine performance on multiple subjects, both within subject and across subject performance [120]. Finally, to test the ability of the CNN to generalize across *unseen* subjects, we explore the performance when testing upon a subject for which no sample data is present within the training dataset – a crucial real-world application.

In summary, the major contributions of this chapter are:

- An improved SSVEP-based system by using the state-of-the-art dry-EEG headset that eliminates rigid procedure and reduce the preparation time.
- A novel end-to-end deep learning CNN architecture to perform the classification of raw dry-EEG SSVEP data without the need for manual pre-processing or feature extraction (the first study to do so with the mean accuracy of 89% when compared to [226]).
- An approach with the ability to generalize to entirely unseen subjects with no additional training, raising the potential for subject-independent BCI applications.

To aid in reproducibility of the results presented in this chapter, all of the associated code has been open-sourced and made available online. The code for the stimuli used in this chapter is available here - https://github.com/nikk-nikaznan/SSVEP_Stimuli/blob/master/onebox_stimuli_RDA.py and the code for the proposed classification of dry-EEG bio-signals in this chapter is available here - https://github.com/nikk-nikaznan/SCU.

3.2 Related Work

A detailed literature review has been presented in Chapter 2. In this section, we focus on a specific subset of comparative works that are closely related to the contribution of this chapter in classifying SSVEP-based dry-EEg bio-signal via a DNN.

In [133], a 32-channel dry-EEG was used on subjects in which they fixated on 11 and 12 Hz SSVEP stimuli during walking trials. The performance and the quality of cortical signals between the wet-EEG and dry-EEG during locomotion were compared. From their experiments, wet-EEG performed better as compared to the dry-EEG by 4% to 10% in accuracy for standing and walking at different speeds, respectively.

The study of foot motor imagery has been carried out in [135] to trigger a lower limb exoskeleton while using the same 20-channel dry-EEG headset we use in this present work. The aim of the study is to employ the quick setup system for asynchronous motor imagery BCI as offered by using the dry-EEG headset.

Deep learning approaches have been used in many different BCI applications, akin to motor imagery [208] as well as the classification of SSVEP signals. In [115], the authors control an

exoskeleton via a visual stimulus generator that had five different frequency LED to control five different behaviours for static and ambulatory experiments. They used eight wet-EEG electrodes to measure the SSVEP signals to compare the classification result between traditional ML against three proposed DNN methods. The data from the stimuli is pre-processed for all approaches with the 3 layers CNN method providing the best accuracy results across both EEG data types.

A five class SSVEP signal problem was classified using both traditional machine learning approaches and deep learning in [226]. The authors analyse data recorded by a traditional wet-EEG dataset via CNN, RNN with LSTM, k-NN, MLP, decision trees and SVM models. Within all the classifiers, the CNN model outperformed others with a mean accuracy of 69.03% and within the traditional classifiers, SVM provided the best overall accuracy.

The authors in [120] introduce EEGNet, a CNN model for wet-EEG data across paradigms. The paper includes four datasets across four different paradigms from different sources with different data sizes. These authors pre-process the data before training on the datasets using different approaches including both a shallow and deep CNN for within subject classification and across subject classification and for all paradigms. Inconclusively, the results demonstrate that different paradigms perform differently for every approach.

In this chapter, we elevate dry-EEG with a beneficial less rigid experimentation procedure to improve the system robustness. We propose using a novel CNN architecture to decode the bio-signals. In contrast to the earlier works mentioned in this section, we explicitly consider an end-to-end approach, without the need for EEG signal pre-processing, to tackle single subject, multiple subject and unseen subject SSVEP-based dry-EEG signal classification challenges.

3.3 Experimental Setup

As discussed in Chapter 1, the archetypal BCI application consists of primary components such as data acquisition and signal processing. In this chapter, we delineate these two critical components in order to establish the setup for the remainder of our work. Firstly, the data acquisition component consists of the hardware to stream and record the bio-signals, and secondly, the external stimuli used during the experimental process. We use dry-EEG technology as the data acquisition hardware and SSVEP paradigm as the stimuli.

We propose the novel SSVEP Convolutional Unit (SCU) as the classifier for comparison against various machine learning techniques from traditional machine learning, to the deep neural network for verification of the quality of the dry-EEG signals. Much of the work from the literature uses a platform such as Matlab for displaying the interface stimuli, recording the signals and decoding the intention [184, 86, 186, 226]. Here, we are using open source programming language, Python, as the primary platform for integrating different key elements of this work.

3.3.1 Dry-EEG Headset

The state-of-the-art dry-EEG data acquisition technique is used in BCI application as an alternative, providing the usage of dry sensors which offers an approach which alleviates the limitation of electrode gel in terms of skin preparation, stable connectivity and comfort during experimentation. However, this technique suffers from inherently reduced signal quality due to the lack of conduit gel, making the classification of such signals significantly more challenging [154].

In this study, we are using the commercially available Quick-20 dry EEG headset (As in Figure 2.1b) from Cognionics Inc. (San Diego, USA) with 20 dry-EEG sensors as highlighted in blue in Figure 3.1. Cognionics Quick-20 system comes with flex sensors to go through the hair to make contact with the scalp and drypad sensors for the skin (for example on the forehead). It is a totally dry-EEG system and employed without the need for skin preparation and just minimum adjustment to ensure all the flex sensors touch the scalp properly throughout the experiment.

This lightweight headset is readily adaptable to different head sizes, and it is both portable and wireless [162, 163, 133, 135]. The sensors come with individual local active shields which helps reduce the need for the rigid experimental condition [31, 163]. The shield protects signals from excessive electrical and movement artefacts even without conducting the protocol in a Faraday cage (a physical shield using conductive material) [160].

This dry-EEG headset provides 19 sensor channels, covering different parts of the brain, and A2, reference and ground (10-20 sensor layout compliant) as shown in Figure 3.1. The 20-channel (Cognionics Inc.) sensor montage [75] has been co-registered with the MNI Colin27 brain (Montreal Neurological Institute Colin 27 atlas). Average sensor locations were obtained by averaging 3-D digitised (ELPOS, Zebris Medical GmbH) electrode locations from ten individuals.

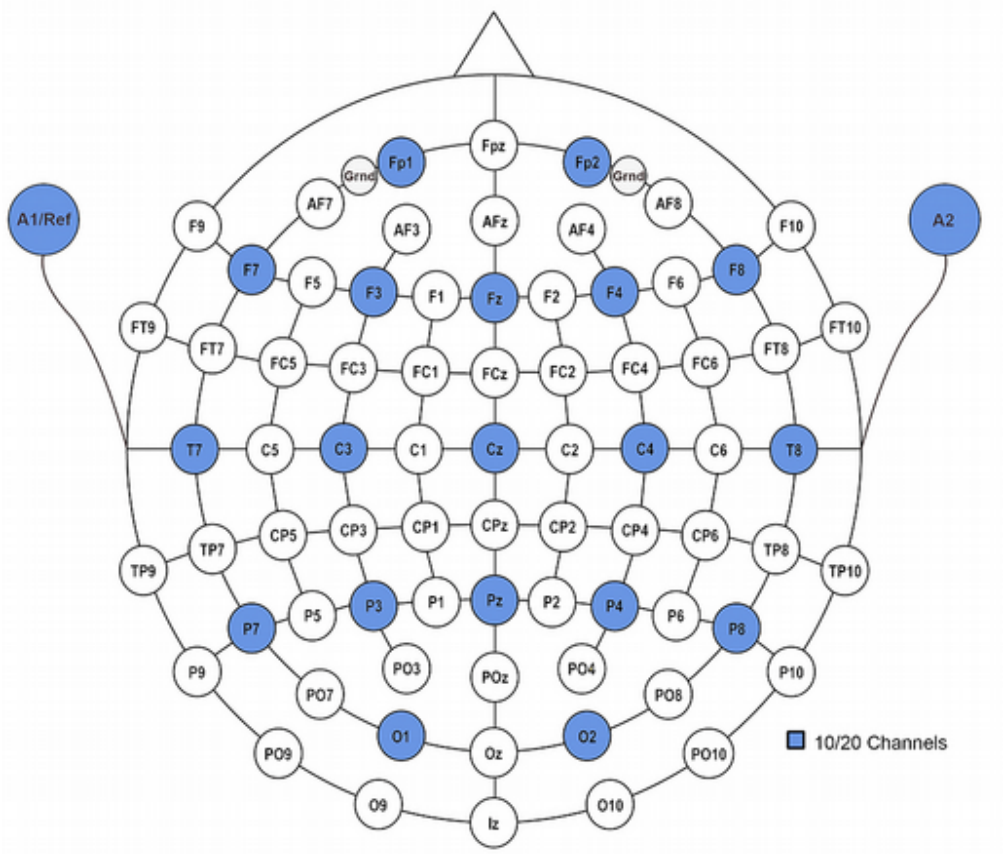


Figure 3.1: Cognionics Inc Quick-20 dry-EEG sensor channels (highlighted in blue)

Electrode labels are assigned based on the nearest neighbour mapping to the standard 10/5 montage. Nas, LPA, and RPA denote nasion and left/right preauricular fiducials [163].

This data acquisition device is a complete system, where it provides the acquisition software to stream and record signals easily using Bluetooth technology. Cognionics data acquisition software is a Windows-based software that is able to connect to the headset and monitor the signals in real-time. The details of the software can be found in Appendix A.

Figure 3.2: Two different patterns are drawn on a PsychoPy window

3.3.2 SSVEP Stimuli

In order to evaluate the quality of dry-EEG bio-signals, we collect dry-EEG signals with SSVEP being used as the neurophysiological response. SSVEP has the feature of frequency tagging, which enables the measurement of neural activity in response to flickering stimuli which the subject is fixated upon. We use Psychopy [184] to present the stimuli for subjects by flickering a square on the monitor with display frequency modulations of 10, 12, 15 and 30 Hz at a time. Psychopy is an open-source package using Python programming language to create stimuli for psychological experiments.

For this SSVEP stimuli, we draw two patterns, a black square as Pattern One (*p1*) and white square as Pattern Two (*p2*) on a window. The frequency desired is determined by how fast the process of changing the stimuli from displaying *p1* to displaying *p2* is performed, as in Figure 3.2. The algorithm used to design the stimuli is detailed in Algorithm 1. We define a frame array for each frequency separately as each frequency has a different array set. These sets determine the pattern's display recurrence via the program, where we we define *p1* as **1** and *p2* as **-1**.

The Frame Array for the frequency modulations used in this experimentation setup is as in Table 3.1. The frequency can be determined by dividing the total length of frame array by the monitor refresh rate. Based on Table 3.1, the length of frame rate of 10 Hz is 6, 12 Hz is 5, 15 Hz is 4 and 30 Hz is 2. Given that the refresh rate of the monitor we used in this thesis is 60 Hz, we can compute that 60 Hz divided by 6 is 10 Hz and so on.

The program draws *p1* first if f_r is **1** and shifts to draw (*p2*) when f_r is **-1**. This processes is repeated for the pre-determined Trial Duration t_d seconds to complete a trial. The window is

Algorithm 1: Window flipping

Data: A frame array [f_r]; `// For example: 10 Hz, f_r = [1, 1, 1, -1, -1, -1]`
Result: SSVEP stimuli

```
Set a window, W as mywin
Set p1 as black square
Set p2 as white square
Set sizes and positions of squares, one for each frequency
while c < n_t do
    flip mywin
    while t < t_d do
        for all n_f ∈ f_r do
            if f_r is 1 then
                draw p1
            end if
            if f_r is -1 then
                draw p2
            end if
            flip mywin
        end for
    end while
    flip mywin to reset W
    wait for w_t
    reset clock for t
    increment c by 1
end while
```

Frequency	**Frame Array (f_r)**
10 Hz	[1, 1, 1, -1, -1, -1]
12 Hz	[1, 1, 1, -1, -1]
15 Hz	[1, 1, -1, -1]
30 Hz	[1, -1]

Table 3.1: Frame arrays set for the frequency modulations

reset to blank during the Wait Duration w_d before the new trial is started. The notation used in the algorithm is presented in Table 3.2.

Symbol	Definition
c	count
n_t	number of trial
t	*clock time*
t_d	trial duration
p_1	pattern one
p_2	pattern two
f_r	frame array
n_f	length of frame array
w_t	wait duration

Table 3.2: Definition and Notation use in Algorthm 1 and Algorithm 2

To stream and save the data in the same platform (Python), we record the signals in real-time from the Cognionics data acquisition software (Section 3.3.1) using the RDA client. RDA or Remote Data Access protocol is one of the integrated streaming protocols provided by the Cognionics data acquisition software to share the signals over the network in real-time.

3.3.3 Data Collection

In this work, we utilise SSVEP as the neuro-physiological response, measured via dry-EEG. The subjects sit in front of a 60Hz refresh rate LCD monitor whilst wearing the dry-EEG headset. We record the data from a range of SSVEP stimuli frequencies; 10, 12, 15 and, 30 Hz [174] using PsychoPy for SSVEP stimuli presentation [184]. The stimuli corresponding to the different flicker frequencies were presented on the primary computer. The setup for this data collection and the communication between the different hardware components is shown in Figure 3.3, with detailed in Appendix A.

During the experiments, we collect data over the parietal and occipital cortex (P7, P3, Pz, P4, P8, O1 and O2) [133], frontal centre (Fz) and A2 reference at 500 Hz sampling rate across four subjects. When having a fixation upon stimulus, the parietal and occipital cortex of the brain are oscillated [235]. The data for subject one (S01) consists of 100 trials of each of the 4 SSVEP

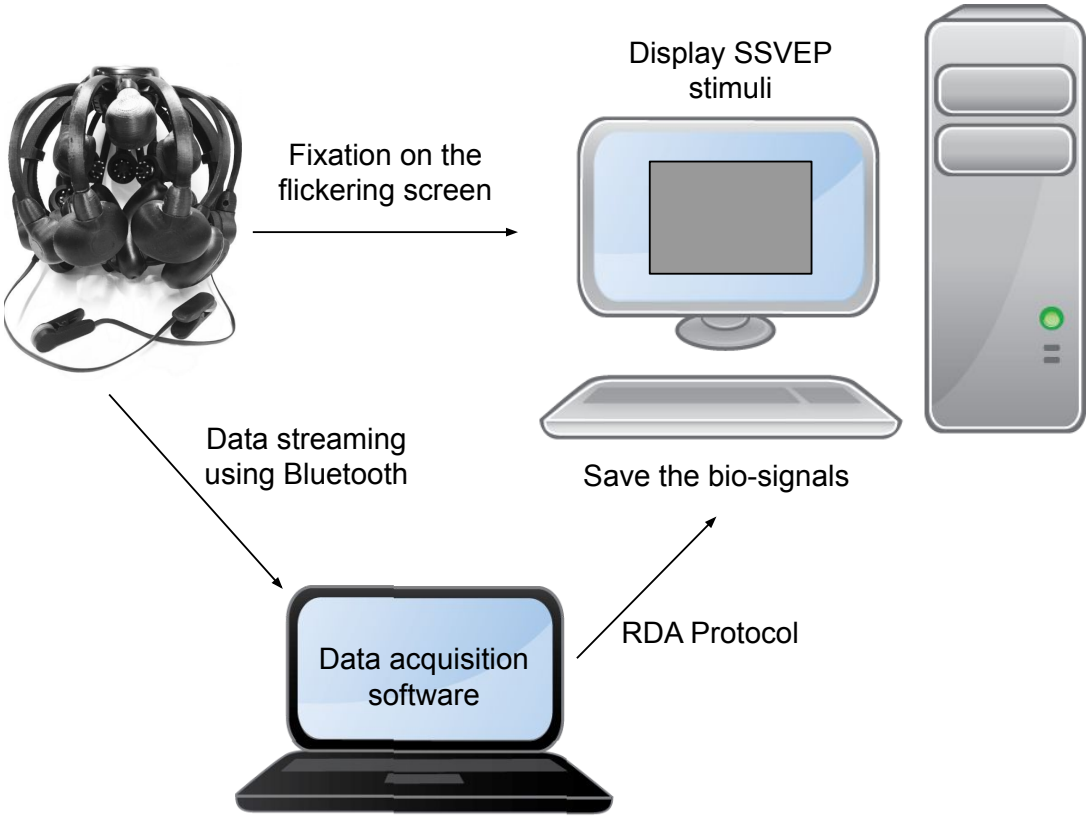
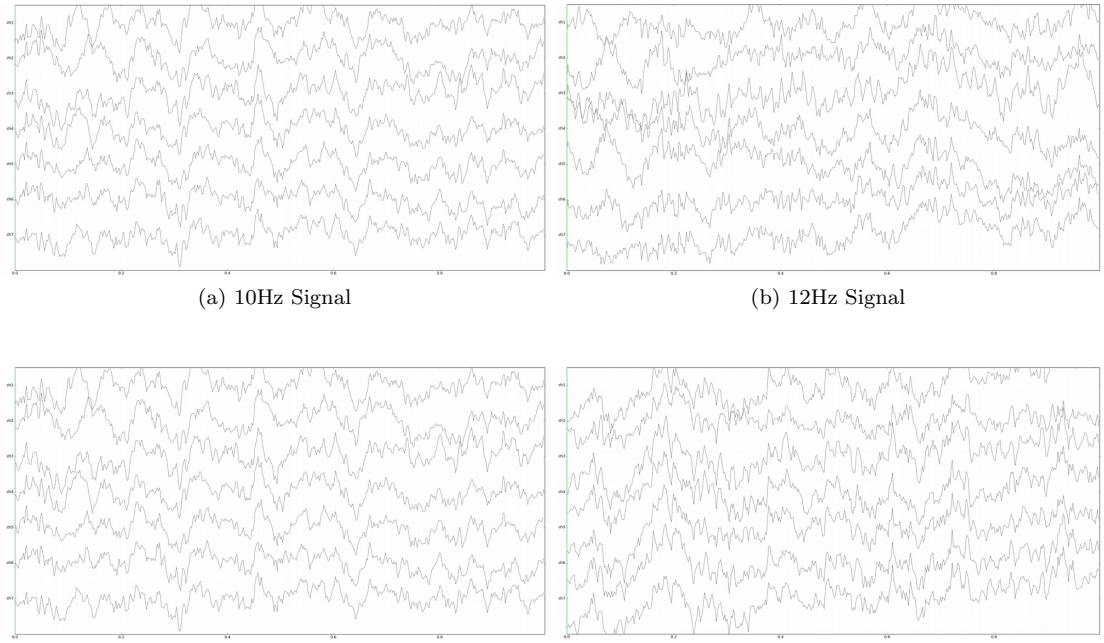


Figure 3.3: Integration set-up between dry-EEG headset and SSVEP stimuli for bio-signals data collection.

classes investigated. Since, collecting EEG signals have been proven to be time-consuming and expensive [53, 253, 145], for the additional three subjects, we only record 20 trials instead. Each trial flickers the LCD screen for a total of three seconds. The data acquisition software is used to monitor and record the signals, as well as providing for real-time measurement of the impedances for the entire duration of the experiment, thus ensuring good quality EEG signals are recorded.

Nevertheless, the primary challenge associated with the classification of dry-EEG signals is the higher noise ratio as compared to the traditional wet-EEG system, owing to the relatively higher impedance values. This noise can be seen in Figure 3.4 which shows the seven distinct dry-EEG data channels across the four SSVEP frequencies we are investigating.



(a) 10Hz Signal

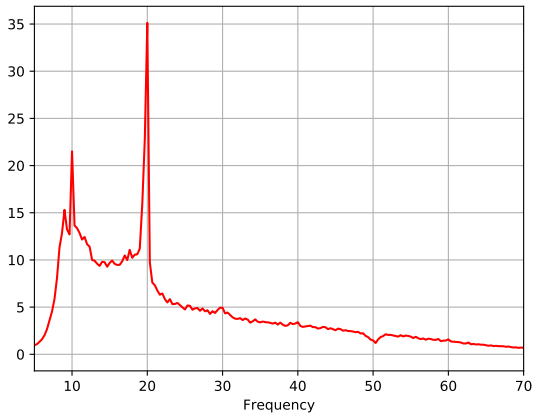
(b) 12Hz Signal

(c) 15Hz Signal

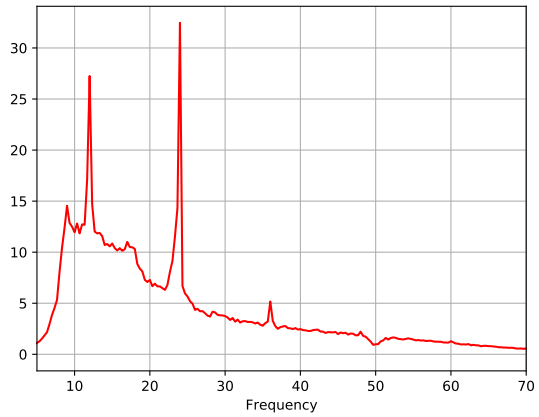
(d) 30Hz Signal

Figure 3.4: Illustrative raw signal data as captured from dry-EEG

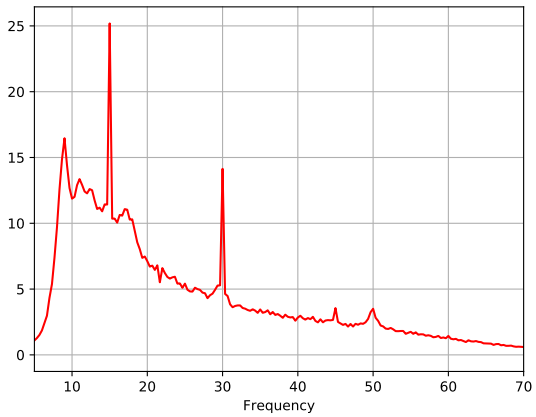
SSVEP data has the phenomenon of frequency tagging, where the primary visual areas in the human cortical oscillates to match the frequency of the fluctuating sinusoidal cycle of the SSVEP stimuli presented to the subject [7]. EEG signals recorded during an SSVEP task will contain the target frequency clearly identifiable in the frequency domain [136]. In order to assure the proposed BCI protocol is reasonable, we use Fast Fourier Transform (FFT) on S01 to decompose the EEG signals into the frequency domain. Figure 3.5 displays the frequency plot of all four frequency classes where they demonstrate the characteristic SSVEP peaks at the target frequency and associated harmonics when taking on average across all samples [194].



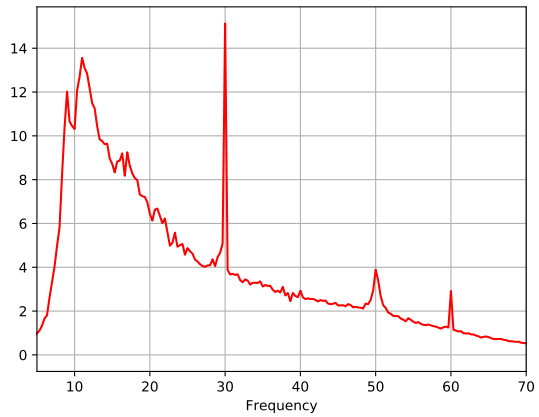
(a) 10Hz Signal



(b) 12Hz Signal



(c) 15Hz Signal



(d) 30Hz Signal

Figure 3.5: FFT plots on four frequency class show the peak frequencies and the harmonics of every classes respectively

3.4 Model Overview and Evaluation Methodology

Signal processing is one of the primary components in the field of BCI and it acts as the translation between the raw EEG cortical signals to a specific desirable decision or application [160]. Traditionally, this requires the use of manual pre-processing and feature extraction stages to transform the data into a format suitable for down-stream prediction tasks. By contrast in this work, we explore the use of a deep convolutional neural network to perform this translation process in an end-to-end fashion.

We explore whether or not a CNN can perform accurate classification of SSVEP target class frequencies on raw dry-EEG data, without the need for manual pre-processing nor feature extraction as found in contemporary work [121]. CNN have demonstrated state-of-the-art results in many image processing tasks, when being used on two dimensional image data [81]. However, there is growing evidence that CNN can be used to process time-series data, when passing a filter over the time dimension, often beating recurrent models designed specifically for such temporal data tasks [18].

3.4.1 SSVEP Convolutional Unit

As EEG data represents time-series data, we make use of a 1D CNN model to classify the dry-EEG data. Our novel CNN model consist of three layers, a convolutional layer, a pooling layer and a final fully-connected layer used to make a prediction. Figure 3.6 demonstrates how the various layers of the CNN process the input bio-signals.

The EEG signal can be considered a 1D input with nine channels (corresponding to the number of sensors). A convolutional filter of size 10 is used to slide over the time dimension of the signals to learn and create the feature representation of signals. The transformed signals are down-sampled to reduce the dimensionality using max pooling in the pooling layer. Max pool works by performing an element-wise maximum over a series of vectors, in this case the various filters of the CNN layer.

The output from the max pooling layer is then flattened and used as input for the fully-connected layer. In this layer, every input additionally has a bias parameter to optimise the

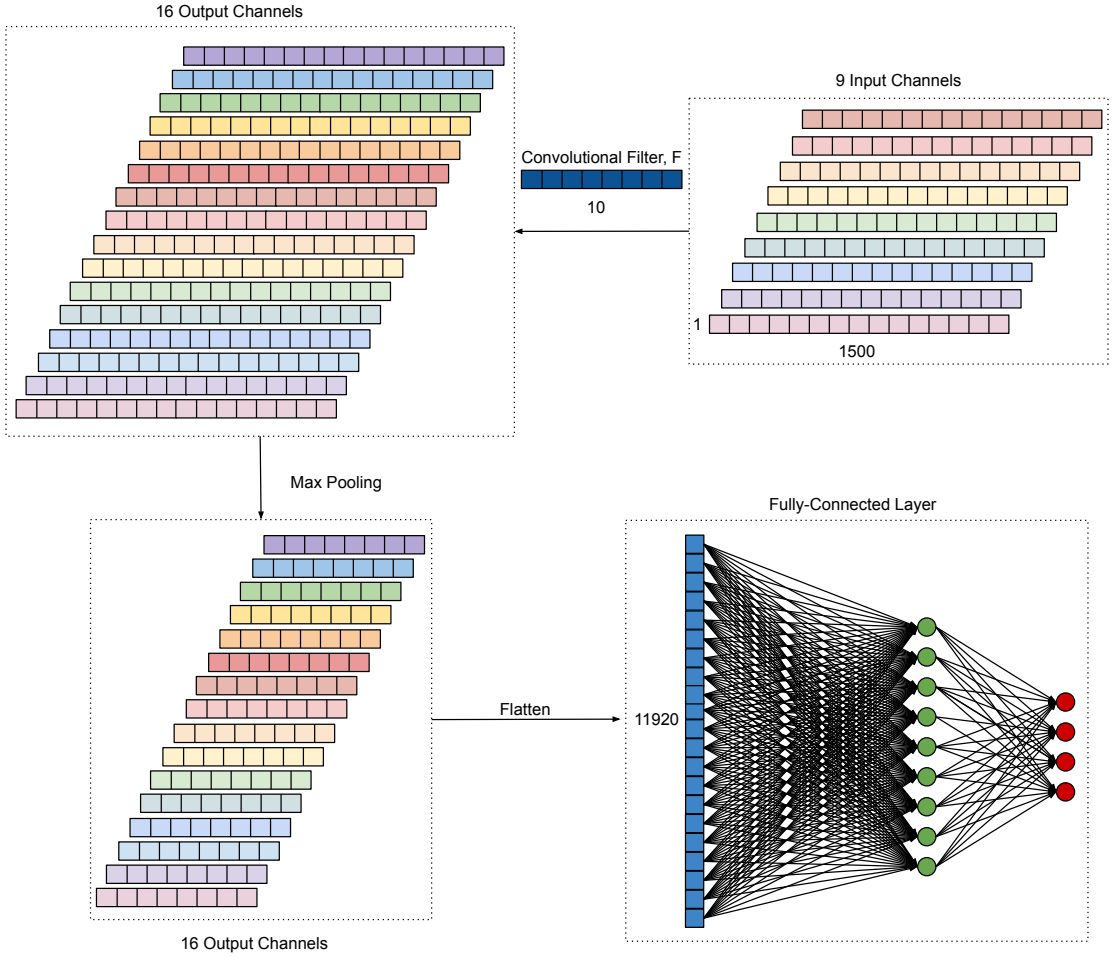


Figure 3.6: 1D Convolutional architecture for dry-EEG signals

learning objectives. For each input sample, this layer will predict a score for all the frequency classes and the maximum score is predicted as the true class of the input signal.

The structure of the CNN used in this work is displayed in Figure 3.7 in which we have our SSVEP Convolutional Unit (SCU) comprising of a triplicate layer of a 1D convolutional layer, batch normalization and max pooling layer operations [81]. These SCU form the common computational building blocks of the CNN architectures used for dry-EEG signal decoding in this section.

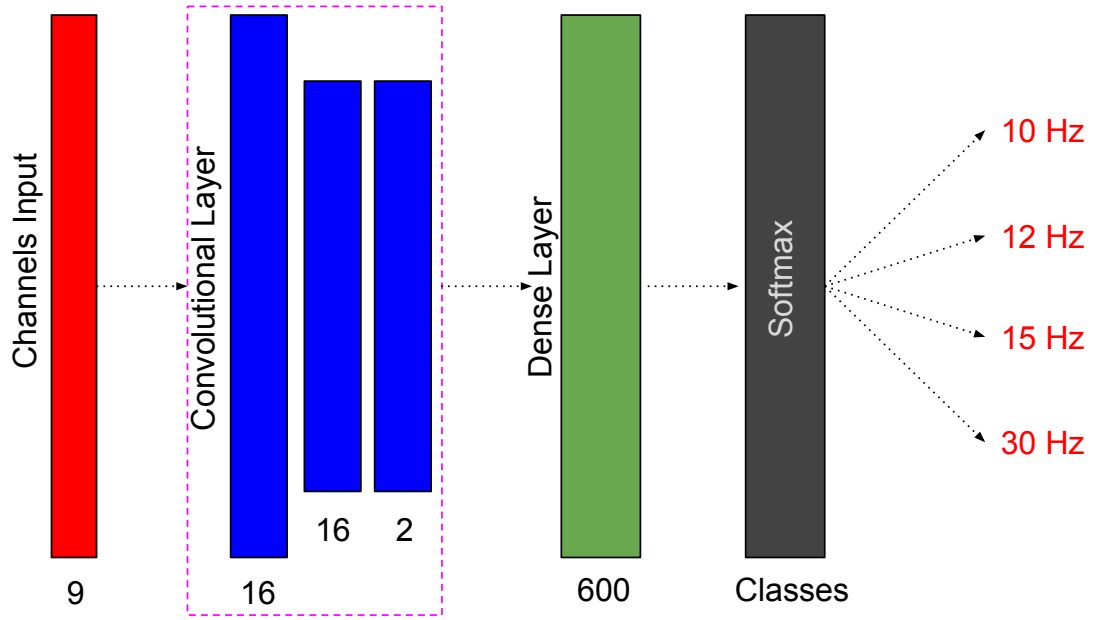


Figure 3.7: Our proposed 1D CNN architecture including our proposed SSVEP Convolutional Unit (SCU, highlighted in pink)

Our CNN architecture has a large initial filter to capture the frequencies we are interested in classifying in the dry-EEG data. We also make use of batch normalization to assist in counter-balancing the noisy EEG data by ensuring all intermediate representations are normalized. Once the data has been transformed via the convolutional filter, the actual classification of the EEG signal is performed via a softmax function (highlighted in black in Figure 3.7) in the final layer. The softmax function takes as input the feature vector x, generated by the CNN $f_{CNN}(y|x)$ and computes the conditional probability of producing the label y as:

$$softmax(y|x) = \frac{\exp\left(f_{CNN}(y|x)\right)}{\sum_{y' \in Y}^{|Y|} \exp\left(f_{CNN}(y'|x)\right)}, \tag{3.1}$$

where Y is the set of all labels in the dataset.

The loss function the model minimised during training is that of Categorical Cross-Entropy (CCE), which will measure the distance between the output distribution of $\hat{y} \in f_{CNN}$ and $y \in Y$ as:

$$CCE(y, \hat{y}) = -\sum_{n=1}^{N}(y_n log(\hat{y}_n)), \tag{3.2}$$

where N is the total number of training samples in the current batch. CCE is a multi-class cross-entropy where cross-entropy is closely linked to KL divergence. KL divergence is used to measure the relative distance between two distributions, in this case how different the predicted distribution is to ground truth distribution.

The model is trained using the ADAM gradient descent algorithm [107], for 100 epochs with a mini-batch size of 32. We also utilise L2 weight decay to help prevent over-fitting by penalising the network for having large weights, meaning that the final objective of our model for optimising is:

$$Loss = CCE(y, \hat{y}) + \lambda ||f_{CNN}(\mathbf{w})||_2^2, \tag{3.3}$$

where $\mathbf{w}$ are the weights of the network and λ is a user controllable scaling parameter, set to 10^{-4} for this work.

In our proposed SCU network, we use a convolution kernel size of size 1×10, a kernel stride of 1, maxpool kernel size of 2, and Rectified Linear Unit (ReLU) as the activation function, as in Table 3.3. Within neural networks architecture, the activation function is used to compute the output value of the learned parameters whilst training the model. ReLU is a common type of non-linear activation function that introduces non-linear transformations into the model, allowing the network to learn complex mapping between the input and the output [81].

3.4.2 Traditional Machine Learning

To validate the effectiveness of our proposed approach, we compare it to traditional classifiers and other deep learning models, which are explored in greater depth in Section 2.3.1. The traditional classifiers used require pre-processing and feature extraction prior to the classification stage. As such, the raw signals are processed via the following steps:- downsampled to 250Hz, referencing to the frontal centre sensor signals (Fz), notch filtered at 50Hz to remove line signal

Parameters	Value
Convolution kernel size	1×10
Kernel Stride	1
MaxPool kernel size	2
Dropout level	0.5
Weight Decay size	0.001
Activation function	ReLU

Table 3.3: Parameters used in the CNN Model

noise and bandpass filtered between 9 to 100 Hz. As a result, pre-processing is used to remove the unwanted signals such as power-line noise, and to focus on the signals between the desirable range [160].

These filtered signals are then utilised as the input for the feature extraction stage. Based on the recent comparative review of [142], we select the Riemannian-based classification approach [20] which utilises covariance matrix and tangent space features which estimate a feature vector in $\mathbb{R}^9$. This approach eliminates the need for feature selection, and by estimating the covariance matrices of the bio-signals and mapping these matrices on a tangent space of the Riemannian manifold.

Based on the result from [226], SVM is the optimal traditional classifier for EEG data. Therefore we use SVM as one of our baseline classifiers with a Gaussian and linear kernel [194]. For further comparison purposes, we also compare with Linear Discriminant Analysis (LDA), one of the frequent choices for EEG analyses [142] and Minimum Distance to Mean (MDM), the frequent classifier for Riemannian-based approach [142, 43].

SVM represents one of the learning algorithms in which it requires the input of a feature vector in order to perform the learning process. SVM is a linear classification method and aims to separate the classes of the input signals upon finding the optimal hyperplane or linear separator as a boundary to separate classes [212].

Linear discriminant function used in LDA represents a hyperplane that is used to distinguish between classes. The true class can be classified by mapping the test feature data from the feature extraction layer into the linear transformation which is called decision function [194].

MDM determine the test class label by computing the distance of the covariance matrices of the test signals to the covariance matrices mapped on the space. The minimum distance is chosen as the class label of that test signal [142, 43].

3.4.3 Recurrent Neural Network

To compare with other leading contemporary neural network approaches, we also compare with several Recurrent Neural Network (RNN) models [134] including vanilla RNN, Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), which have also been assessed for EEG classification in a previous study [226].

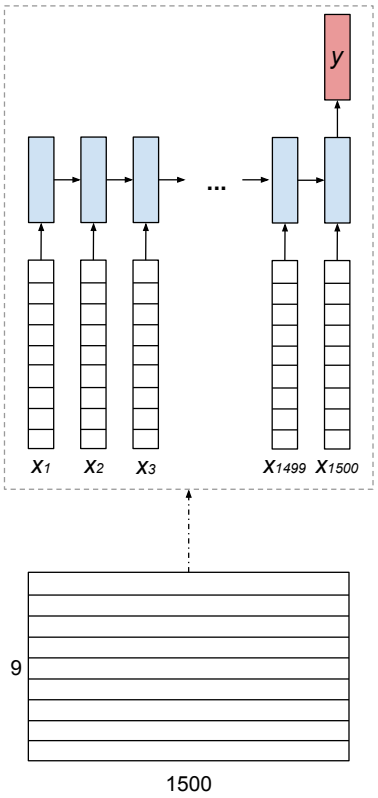


Figure 3.8: RNN sequences used to classify dry-EEG bio-signals

Instead of having convolutional windows like a CNN, RNN computes the signals in time domain where the model predicts the output of current state conditioned upon the previous state for each time-step [27]. Figure 3.8 shows the sequences for an RNN to classify the dry-EEG signals. Each sequence of the time-step is processed by a hidden state where the loss and backpropagation are done at each time-step.

As the network computes the input data at every time-step, a vanilla RNN struggles to remember the whole sequence, causing the loss of important characteristics of the input, especially for the long input sequences. Hence, LSTM [93] and GRU [47] were developed which introduce gates to regulate the information relevant in helping to decide what is important to remember.

For our experiments, we use three different RNN architectures; vanilla, LSTM and GRU, with common key hyperparameters across all networks (detailed in Table 3.4).

Parameters	**Value**
Learning rate	0.0001
RNN units size	2
Dropout level	0.001
Weight Decay size	0.001
RNN layers	2

Table 3.4: Parameters used in the RNN Model

3.5 Results and Discussion

In this section, we present detailed experimental evaluation demonstrating the ability of our approach to accurately classify SSVEP signals in dry-EEG data. Due to the limited data, we divide our data into train and test set only. To avoid accuracy variance problem, we use 10-fold cross-validation technique to train the model on different train sets and results are presented as the mean of accuracy over the test sets. We choose the same seed parameter and use the same partitioning scheme to ensure the fair comparison across all machine learning methods.

All the models were initially optimised via grid-search for hyperparameters selection over k-fold validation. All the presented results are produced based on the best set of hyperparameters

from this process. We present the mean and standard deviation over 10-fold validation set for all the classification results on different machine learning techniques. The dataset and experimental setup are detailed previously in Section 3.3. One initial interesting aspect to note before assessing the predictive performance of the approaches to consider the time taken to train the models [3] where the CNN is a much more efficient model to train.

3.5.1 SSVEP Classification on Subject S01

The results for the classification of the data of a single subject (S01) are presented in Table 3.5. The table highlights the accuracy of our proposed CNN approach against all the baselines discussed in Section 3.4.2 and Section 3.4.3.

Method	**Accuracy**	
	Pre-processing	*Without*
SCU	0.99±0.01	0.96±0.02
Vanilla RNN	0.71±0.21	0.91±0.05
LSTM	0.78±0.24	0.57±0.30
GRU	0.97±0.08	0.90±0.06
$SVM_{Gaussian}$	0.92±0.01	0.86±0.01
SVM_{Linear}	0.94±0.01	0.83 ±0.02
MDM	0.76±0.01	0.80±0.02
LDA	0.90±0.01	0.83±0.01

Table 3.5: Mean accuracy with standard deviation over 10-fold cross validation for subject, S01

The results for the traditional approaches, are presented with and without pre-processing. Without pre-processing, we only perform feature extraction before classification with the traditional baseline approaches. Overall, the results show that, even without any pre-processing of the data, our SCU approach demonstrates superior performance over the baselines. The confusion matrix obtained from the classification of S01 using the SCU is presented in Figure 3.9, which shows very strong accuracy across all classes.

[3] The average training time taken for vanilla RNN is 600 minutes, GRU 524 minutes, LSTM 619 minutes and 4 minutes for CNN on Nvidia GeForce GTX 1060 GPU

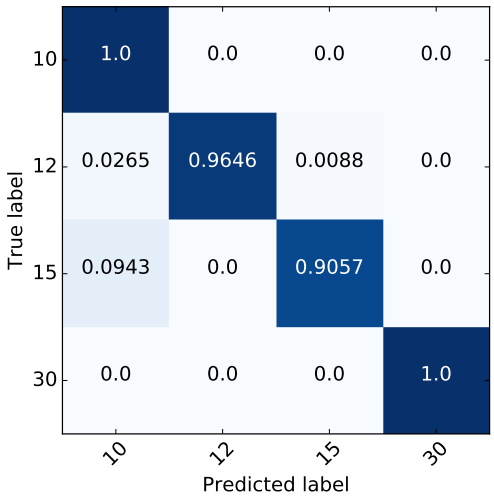


Figure 3.9: SCU classifying on single subject (S01) for frequency classes (Hz), highlighting the per-class statistical accuracy (Maximal result being accuracy = 1.0 in the matrix diagonals)

3.5.2 SSVEP Classification on Multiple Subject

The second set results, presented in Table 3.6, demonstrate the classification performance across three subjects {S01, S02, S03}, where a new classification model is trained for each subject. Due to the known impracticalities of collecting large amounts of data per subject [142], here we reduce the number of SSVEP presentation sessions (trials) per subject for each class to only 20. Based on the results presented in the previous result of Section 3.5.1, and for the sake of brevity, for the remaining experiments we only consider the highest performing classification approaches across traditional and deep model approaches without any pre-processing (SCU and SVM; Table 3.5).

The results highlight, that even with a reduced quantity of data available, the SCU approach still significantly outperforms the SVM across all subjects. This result highlights the applicability of the proposed SCU approach for BCI applications, where data quantity is often relatively limited [142].

Method	**S01**	**S02**	**S03**	textbfMean
SCU	0.91±0.08	0.92±0.11	0.85±0.10	**0.89±0.03**
$SVM_{Gaussian}$	0.59±0.08	0.68±0.08	0.67±0.10	0.65±0.04
SVM_{Linear}	0.76±0.05	0.68±0.07	0.58±0.10	0.67±0.07

Table 3.6: Mean accuracy with standard deviation over 10-fold cross validation for each of the three subjects, mean results across subjects are also presented

3.5.3 SSVEP Classification Across Subjects

To assess the ability of a single SCU model to classify a dataset comprising data from all of the subjects {S01, S02, S03}, we classify all the signals from the three subjects together instead of performing individual classification.

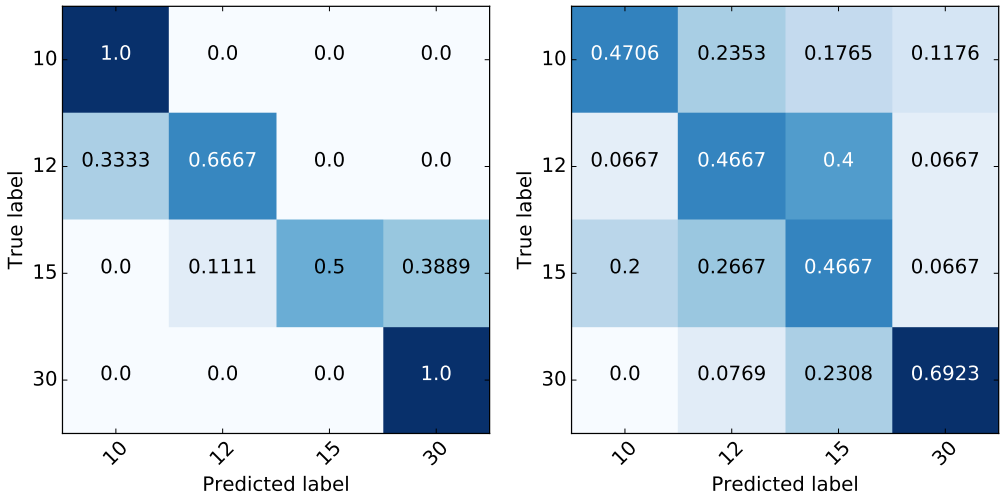
Method	**Accuracy**
CNN	**0.78±0.10**
$SVM_{Gaussian}$	0.51±0.06
SVM_{Linear}	0.50±0.06

Table 3.7: Test Accuracy across subjects

Having a single model trained on EEG data from multiple subjects is known to be challenging [56], potentially due to biological differences between subjects and the variability of the EEG recording process. However, the results presented in Table 3.7 show that our SCU model is able to significantly outperform the SVM based approaches when performing classification across subjects. This can be further seen in figure 3.10, showing better performance across classes for the SCU.

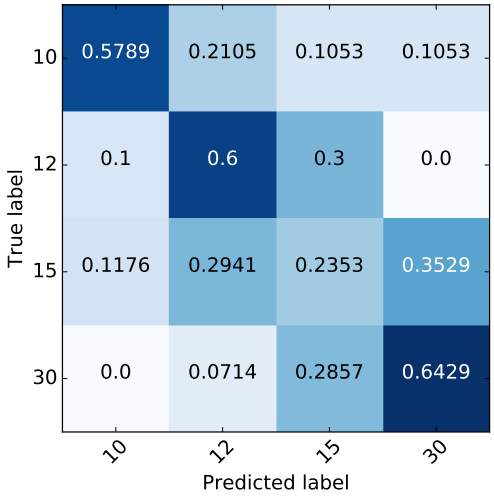
3.5.4 Generalisation Capability to an Unseen Subject

A strongly desirable quality for any model performing the classification of EEG is that of unseen subject generalisation - whereby the model is able to correctly classify data from a subject whose data is absent *a priori* model training. To test this on our SCU model, we introduce the data of the *unseen* subject S04. We then attempt to classify these data using a model which



(a) CNN classifying across subjects.

(b) $\text{SVM}_{Gaussian}$ classifying across subjects.



(c) SVM_{Linear} classifying across subjects.

Figure 3.10: Confusion matrices for the SSVEP classification across subjects for frequency classes (Hz), highlighting the per-class statistical accuracy (Maximal result being accuracy = 1.0 in the matrix diagonals).

was trained only on the data of the other three subjects, {S01, S02, S03}. Using the same CNN architecture for this task as depicted in Figure 3.7, we only achieve an accuracy of 0.59 on S04

without any additional training. We also attempt to classify the new test subject using SVM, however the SVM only displays random classification performance (≈ 0.25 accuracy; i.e 1/4 for 4 classes).

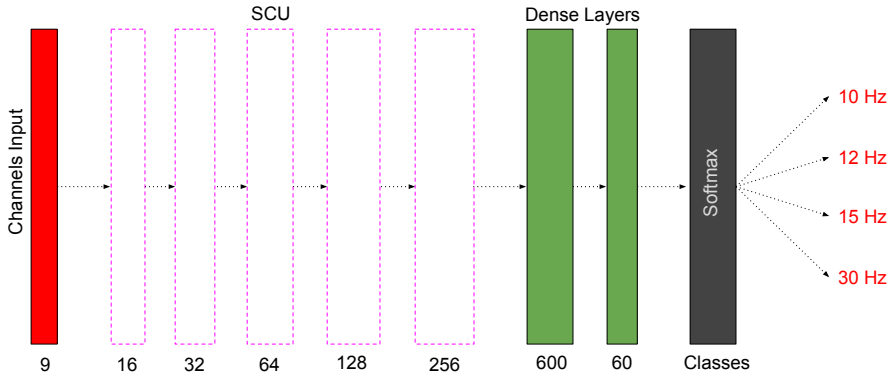


Figure 3.11: Our deeper CNN architecture for *unseen* subjects

To overcome this performance issue, we explore a deeper architectural network variant as deeper networks have been shown to learn more complex features in order to determine the correlation between subjects [81]. Figure 3.11 illustrates the deeper architecture where empirically, we repeat our SCU blocks (each dashed pink box represents a SCU block) to a maximum number of five. This deeper architecture, introduced to classify subject S04 data, demonstrates a substantially better classification accuracy of 0.69, perhaps suggesting that a deeper model is required to perform the unseen subject generalisation task. The confusion matrix for this result is presented in Figure 3.12. This figure demonstrates that the CNN has varying performance across the different classes, with the 30Hz signal being the best performing for this extended CNN model.

3.6 Conclusion

In this chapter, we have introduced SSVEP-based system by using the state-of-the-art dry-EEG headset that eliminates rigid procedure and reduce the preparation time. We proposed a novel deep convolutional neural network architectures constructed around a common computational building block, for the classification of raw dry-EEG SSVEP data, the first such study to do so. We evaluate the performance of our model on SSVEP data recorded from four

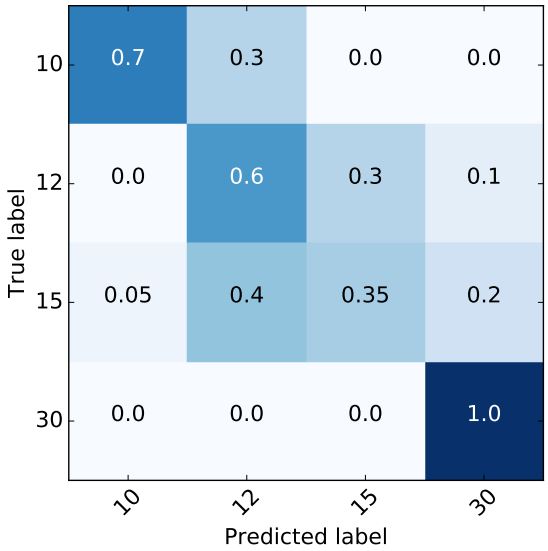


Figure 3.12: Confusion matrices for the SSVEP classification on unseen subject (S04) for frequency classes (Hz) by using Deep CNN highlighting the per-class statistical accuracy (Maximal result being accuracy = 1.0 in the matrix diagonals)

subjects using the noise-prone dry-EEG methodology. As compared with current state-of-the-art methods [226], our approach requires no pre-processing of the data, demonstrates higher overall classification accuracy across subjects and generalizes significantly better to entirely unseen test subjects. These key results demonstrate that CNN based approaches should become the new benchmark method for SSVEP dry-EEG classification.

Although this work proves the ability of CNN to accurately classify dry-EEG bio-signal, there are some limitations within this scope. We only use simple stimuli with a single box stimulus at a time which make the system less applicable with the lack of ability in controlling multiple commands at a time. The advantage of using one single box stimulus at a time is that it can lead to having less noise interference in the recorded signals and always knowing the correct ground truth (i.e subject has no other stimulus present to distract them). The results presented here are based only on four subjects EEG signals, which need some expansion for further justification of the proposed BCI system. Finally, the classification performance of the proposed SCU has only be tested for offline setting which should be tested further to show the ability to classify in a real-time application.

Future work based on the research presented in this chapter could involve larger datasets taken from a larger set of subjects in order to further study the classification and generalisation performance across subjects. Additionally, more complex SSVEP stimuli with online decoding capability would improve the application of this approach to the field of BCI. Finally, the combination of the CNN and RNN models may also offer a way to increase overall performance, perhaps at the cost of longer training times.

Chapter 4

SSVEP-based Variable Natural Environment Stimuli

This chapter addresses the challenge of humanoid robot teleoperation in a natural indoor environment via a Brain-Computer Interface (BCI). We leverage deep Convolutional Neural Network (CNN) based image and signal understanding to facilitate both real-time object detection and dry-Electroencephalography (EEG) based human cortical brain bio-signals decoding.

We employ recent advances in dry-EEG technology, as in Chapter 3, to stream and collect the cortical waveforms from subjects while they fixate on variable SSVEP stimuli generated directly from the environment the robot is navigating. Uniquely, instead of using the empty square traditional stimuli as in previous chapter, in this chapter we use the variable stimuli in term of pixel pattern and square size in natural environment. We utilise the real-time video streamed via the on-board robot camera as visual input for SSVEP, where the CNN detected natural scene objects are altered and flickered with differing frequencies (10Hz, 12Hz and 15Hz). These stimuli are not akin to traditional stimuli - as both the dimensions of the flicker regions and their on-screen position changes depending on the scene objects detected. We stream and collect the cortical waveforms from subjects whilst they fixate on the designed stimuli generated directly from the environment the robot is navigating. On-screen object selection via such a dry-EEG enabled SSVEP methodology, facilitates the on-line decoding of human cortical brain

signals, via a specialised secondary CNN, directly into teleoperation robot commands (approach object, move in a specific direction: right, left or back). This SSVEP decoding model is trained via *a priori* offline experimental data in which very similar visual input is present for all subjects.

In addition, taking advantage of having recorded the streamed video, we collect a total of additional nine subjects for use as a offline dataset for further study. As the stimuli is more complex compare to Chapter 3, we run the SSVEP classification experiment to ensure the quality of the signals is sufficient. We compare against different classification methods, and can thus be seen as strengthening the contribution of the previous chapter as a similar model also works well on a more complex task across a larger number of subjects.

Some parts of the work presented in this chapter have been published in the following peer-reviewed publication:

> *Nik Khadijah Nik Aznan, Jason D Connolly, Noura Al Moubayed, and Toby P Breckon. Using Variable Natural Environment Brain-Computer Interface Stimuli for Real-Time Humanoid Robot Navigation. In* International Conference Robotics and Automation*, pages 4889–4895. IEEE, 2019*

4.1 Introduction

Teleoperation or telepresence is a field within robotics which has been widely utilised for numerous applications. It allows humans to remotely control robots, either whilst being present within the same location, or remotely over the internet [26]. In this work, a humanoid robot is used as a teleoperational remote control interface, allowing a human to navigate the robot via the use of BCI-based cortical brain bio-signals [194]. This application can be used widely, for example by severely disabled people as an alternative communication platform with the robot without requiring any actual physical movement from the patient [210].

In order to gather the cortical signals from a human test subject, a non-invasive dry-EEG is used. EEG is a technique where electrodes or sensors are placed on the scalp to capture electrical activity of the brain, without the need to implant them directly into the brain, such as invasive microelectrode arrays [160]. We utilise the Cognionics Quick-20 dry-EEG Headset

that requires no conductive gel and has the additional benefit of being a wireless device; as compared to traditional wet-EEG [133, 163]. By contrast, wet-EEG requires the cumbersome application of conductive gel, the use of a restrictive Faraday cage enclosure that prohibits real-world application and semi-invasive blunted needles that scratch the skin surface in order to lower the impedance values and hence improve the connectivity between the electrodes and the scalp [139]. Dry-EEG is an alternative approach used to improve the usability of EEG within a BCI context via the elimination of these wet-EEG factors [135].

This chapter explores the creation of a BCI-based application to accurately navigate a humanoid robot in an open environment via the above noted dry-EEG headset. We make use of NAO, a humanoid robot [83] which is equipped with cameras and programmable movement and behavioural features such that different commands can be interpreted to navigate the robot to move toward the object of participant visual and cognitive interest [258]. To develop this effectively, we employ the features available from the Nao platform such as streaming the real-time video from the robot as visual input for the SSVEP stimuli. As explored in the previous chapter, SSVEP is a type of stimulus-evoked neurophysiological response induced simply via subject fixation (or even just via peripheral attention) on visual stimuli and requires almost no *a priori* user training [137, 45, 169]. Different to the eye-tracking that requires gaze control, SSVEP responses are a covert condition response where the subject does not need to control their gaze [170]. The human cortical signals in the primary visual areas oscillate when visually evoked via these stimuli by a continuously fluctuating sinusoidal cycle [7, 151].

In this work, we propose a novel variable dry-EEG enabled BCI stimuli for robot navigation utilising a pre-trained object detection convolutional neural network. We perform object detection in real-time derived from the incoming video stream from the robot camera. One of the primary contributions of this work is to make the SSVEP stimuli more natural to the user, as the stimuli (or in our case, objects) is presented in the context of the real-world scene the robot is currently navigating. Unlike previous stimuli [133, 38, 137], in this work the size of each SSVEP flicker region depends on the physical dimensions of the object detected. The detected object pixel regions are flickered at differing on screen frequencies (10, 12, 15 Hz) and the decoded EEG signals are used to navigate the robot to walk toward objects based on the objects selected by the subject (user) via the SSVEP interface.

To perform the dry-EEG signal decoding we use the SCU CNN-based architecture, detailed in the previous chapter, to differentiate between EEG signals by extracting unique features across

multiple layers of convolutional transformation optimised over a set of training data [81]. This model is used to classify real-time dry-EEG signals before sending the decoded command to navigate the robot towards the scene object that the subject has selected [208].

Following standard practice in the BCI literature, we evaluate the performance of our work by testing the classifier model on real-time humanoid navigation via classification accuracy and Information Transfer Rate (ITR) as performance metrics - the latter representing a quantitative measure of the speed of BCI information transfer [194].

In summary, the major contributions of this chapter are:

- Use of novel variable position and size SSVEP BCI stimuli, based on using object detection pixel regions identified in real-time, within the live video stream from a teleoperated humanoid robot traversing a real-world natural environment.
- An offline dry-EEG enabled SSVEP BCI signal decoding (classification) result achieving mean accuracy of 84% with the use of variable stimuli size and on-screen stimuli positioning (the first such study to accomplish this).
- Demonstrable real-time BCI teleoperation of a humanoid robot, based on the use of naturally occurring in-scene stimuli, with a peak mean accuracy of 85% and ITR of 15.2 bits per minute (bpm) when evaluated over multiple test subjects (teleoperation users).

Our proposed novel variable stimuli is available here - https://github.com/nikk-nikaznan/SSVEP_Stimuli.

4.2 Related Work

The detailed review on related literature has been presented in Chapter 2. In this section, we focus on the studies utilising humanoid robots with EEG signals within SSVEP-based BCI applications.

The work of [258] proposed the use of behaviour-based SSVEP to control a telepresence humanoid robot to walk in a cluttered environment, with the tasking of approaching and picking

up a target. They controlled the robot by classifying 4 sets of movements with a total of fourteen behaviours of the robot. One visual stimuli is used to select the behaviour set and the remainder are used to encode the behaviours. The user interface of the system consists of five fixed stimuli symbols (five frequency values), a display for a live video feedback and a display for the current posture of the robot. The task completed with an average success rate of 88%, an average response time of 3.48 s and an average ITR of 27.3 bpm.

Similar research has been carried out using SSVEP stimuli to control robot-like behaviour in [137] and [38] in which these authors again used fixed size and position stimuli symbols with differing frequencies that indicate different directions for the robot to move toward. In [137] the authors controlled a mobile robot by using 3 different SSVEP frequencies by moving forward or turning to the left or right in order to avoid the obstacles. The stimuli in [38] consisted of four fixed flickering boxes where each frequency was used to command a mobile robot (forward, backward, turn counter-clockwise/clockwise) to navigate the robot through a maze path.

There are two notable studies that have integrated object detection and recognition [210, 227]. In [210], the authors used seven different frequencies to navigate a mobile robot to a storage rack to grasp an object and delivered into a dustbin with an average mean accuracy of 89.4%. The approach employed an AdaBoost algorithm with Haar features to recognise three objects on the rack for subjects to choose. However, the recognised objects were not flickered as stimuli - instead, there were separate fixed stimuli designed with three different frequencies corresponding to each object.

The authors in [227] used SSVEP with a hybrid-mask feature in which a 3D textured model was rendered and flickered on certain scene objects. In this case, three similar cans which are recognised offline. Subjects for this study teleoperated a humanoid robot HRP-2 (located in Japan whilst the subjects were physically located in Italy) to control the robot to firstly, grasp a can from a table, navigated the robot to a second table where the robot is required to drop the can on a marked target.

In this chapter, by taking advantage of the on-board camera on our humanoid robot and the high-performance scene object detection model of [138], we instead use variable BCI stimuli, embedded within the scene video feed. This is achieved by flickering the flexible size detected object pixel regions with differing SSVEP frequencies. This occurs in the real-time as the humanoid robot navigates a natural indoor environment. In contrast to earlier work [258, 137, 38, 210, 227], our stimuli vary both in terms of pixel pattern, size and on-screen position in-conjunction with the changing nature of the environment the robot is navigating through.

4.3 Variable BCI Stimuli

In order to translate the cortical signals, we use SSVEP as the neurophysiological brain response for our BCI experiment. The stimuli are embedded into a real-time video streaming from the on-board robot camera (RGB colour, resolution: 1280×960). Based on pre-trained object detection generated by a pre-trained network, we flicker the on-screen display frequency of objects by rendering black/white polygon boxes on top of the objects with display frequency modulations of 10, 12 and 15 Hz [151].

4.3.1 Single Shot MultiBox Object Detector

In this chapter, we propose a novel variable dry-EEG enabled BCI stimuli for robot navigation utilising a pre-trained object detection convolutional neural network. Our key idea is to make the SSVEP stimuli more natural to the user, as the stimuli (or in our case, objects) is presented in the context of the real-world scene the robot is currently navigating. In this work, the size of each SSVEP flicker region depends on the physical dimensions of the object detected. The detected object pixel regions are flickered at differing on screen frequencies. To perform the object detection, we employ the pre-trained Single Shot (MultiBox Object) Detector (SSD) CNN [138].

SSD is a CNN-based object detection model designed for real-time prediction. This model utilise convolutional layers of varying sizes targeting different object size in a scene. The variety sizes provides multi-scale feature maps as when the spatial dimension of the model is reduced, the resolution of the feature maps is decreased as well. The high resolution feature maps focus on detecting small object whilst the low resolution one is for detecting the large object. During training, the model predicts multiple prediction for each pixel cell where each prediction has a boundary box and confidence scores. Gradient descent is used to optimise the prediction to the ground truth. SSD was originally trained by using 12 objects from the COCO dataset; aeroplane, bicycle, bird, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, potted plant, sheep, sofa, train and tv monitor [131].

We exploit the bounding box and label prediction returned from the network and present the detected objects as SSVEP stimuli via Psychopy [184]. We employ the use of a pre-recorded video for offline experiment and as calibration stage, we use on-board video for streaming from the natural environment for real-time experiment. Both videos are used as stimuli for offline and online experiments respectively where the videos are the input for SSD object detection model.

4.3.2 Stimuli Setup

The SSD network uses OpenCV to display the video frame, and to draw and display the bounding box [28] and Psychopy for displaying the stimuli. The detailed process of getting a frame to display as a stimulus can be find in Appendix B. In this chapter experimental setup, we use five objects which are chair, person, dog, car and plotted plant as objects in the natural experimental scene. By using a humanoid robot, NAO, we stream the natural surrounding using the on-board NAO's camera via Application Programming Interface (API), Naoqi provided by Aldebaran software that comes with the humanoid robot.

We assign up to two objects and an arrow per frequency. The objects that are appointed to the same frequency will never appear in the same experimental scene. We set two different stimuli, one for the calibration stage and one for the real-time navigation experiment. The calibration stage is conducted before the real-time navigation experiment where we use pre-recorded video streamed from the humanoid robot to be the stimuli. The bio-signals gathered from this stage are used to train the SCU SSVEP classification network. For the real-time navigation experiment, the video from the on-board camera from the humanoid robot is streamed in real-time whilst the robot is moving around in the natural environment.

The teleoperation interface display alternates between this detected object flickering and the navigational arrow flickering, alternately as illustrated in Figure 4.1. The additional use of the navigational arrow stimuli enables the subject to navigate the robot when there is no new object detected within the scene, for example, when the robot is too close to the previously subject (user) selected object.

4.4 Methodology

In this section, we present the three other primary experimental components apart from variable BCI stimuli, those being the calibration stage, EEG signal classification and robot navigation. The overall setup and data flow of the experiment is shown in Figure 4.2.

We use the on-board camera to stream video from the natural environment to a monitor display in front of the BCI subject (user). Using the CNN-based object detection model of [138],

Figure 4.1: Variable SSVEP stimuli based on object detection flickering (left column) to navigate the robot towards the object and navigational arrow flickering (right column) to move the robot facing a new environmental scene.

detected scene objects are identified and flickered with a unique on-screen SSVEP frequency (from set: 10Hz, 12Hz, 15Hz). EEG signals from the subject are streamed using the dry-EEG headset whilst they fixate on a flickering on-screen scene object. A SCU-based CNN pre-trained on an *a priori* offline dataset is then used to infer the class of the EEG signals in real-time. This prediction is used to navigate the robot towards the corresponding scene object the subject is fixated upon.

4.4.1 Calibration Stage

We use the pre-recorded offline video to collect some offline data to be used to train the network for SSVEP classification. This is known as calibration stage within the area of BCI applications [141, 2, 56, 248]. This stage is conducted right before the real-time navigation experiment to ensure the same session condition is secured to avoid any inconsistency. We

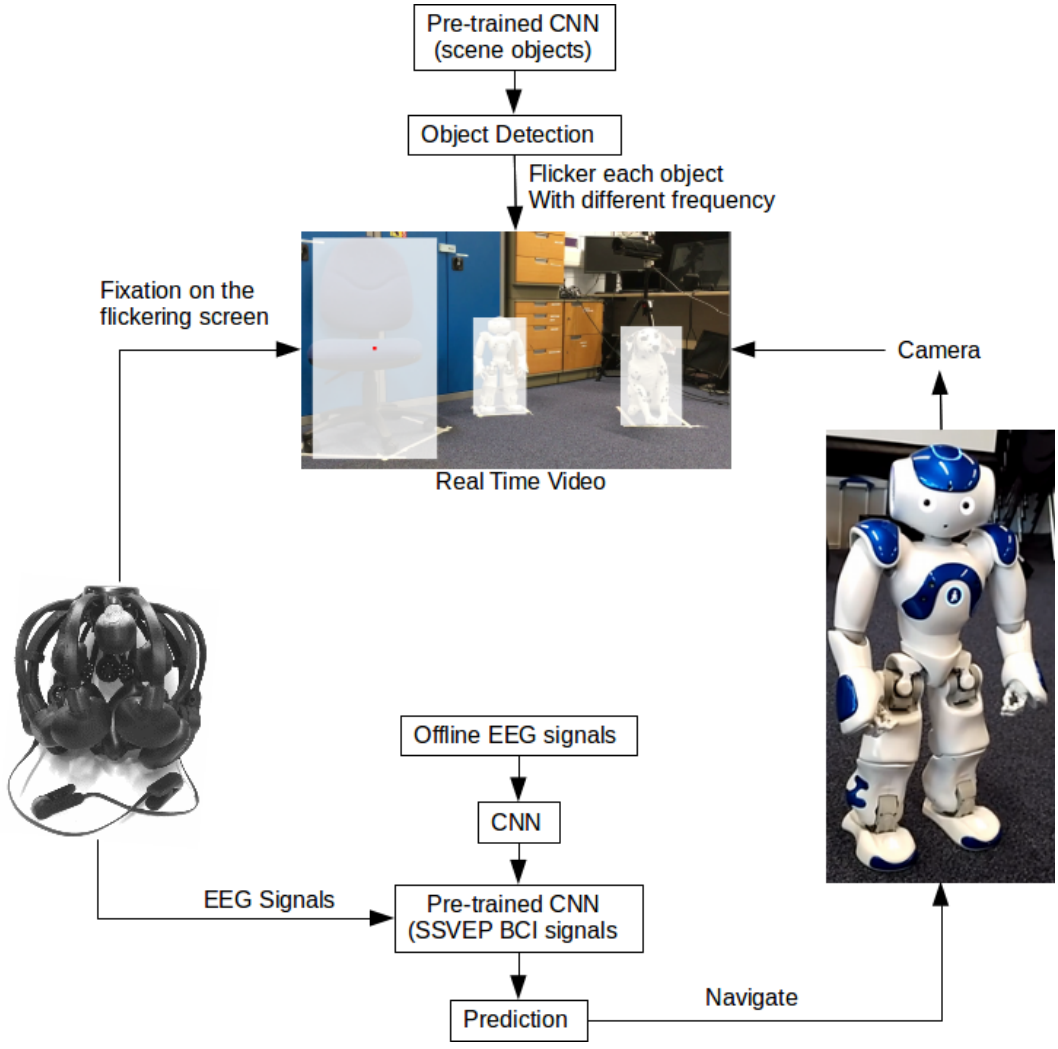


Figure 4.2: Overview of the experimental approach proposed.

display object stimuli followed by arrow stimuli one after another during the trial. We are using pre-recorded video from the on-board camera on the humanoid robot as the input for the visual stimuli pipeline. This calibration stage was performed for each subject, where 60 samples per class are collected. We use this offline dataset as input to the CNN to train the network specifically to classify the bio-signals of the specific subject. The pre-trained weights are saved and used to decode the test signals during the real-time navigation experiment.

4.4.2 SSVEP Classification

To decode the dry-EEG signals efficiently in order to ensure effective teleoperation of the robot, we use our deep CNN architecture previously introduced in Chapter 3, for signal to object/motion label classification.

During the offline experiments, subjects fixate to one of the flickering stimuli. The cortical brain signals from each subject are collected for 40 experimental trials per SSVEP class to form the offline *a priori* training sets or training the CNN model per subject (offline calibration).

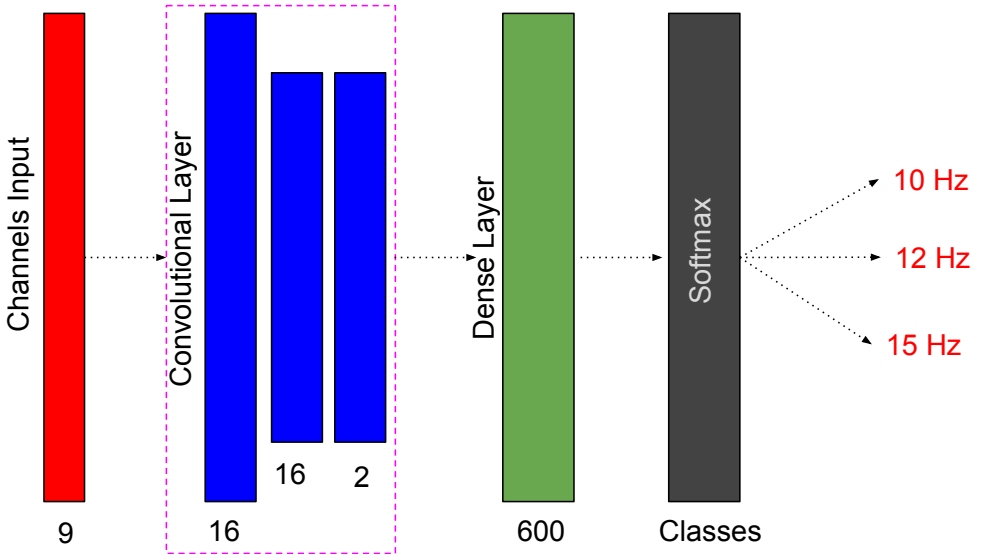


Figure 4.3: The 1D CNN architecture used to classify the EEG signals for both offline dataset and real-time experiments (SCU, highlighted in pink).

We train a SSVEP Convolutional Unit (SCU) CNN architecture, comprising of a 1D convolutional layer, batch normalization and max pooling (as detailed in Figure 4.3) by using the offline priori experimental datasets. We first bandpass filter the incoming sigmals between 9 to 100 Hz in order to reduce undesired high or low frequencies that are not of interest in this work. The filtered signals, which consist of nine input channels, are transformed by using a large initial convolutional filter to capture the frequencies we are interested in classifying within dry-EEG signals. The SCU CNN model is trained using backpropagation with stochastic gradient descent [121].

For this training, the key hyperparameters, initially chosen via a grid-search over a validation set, are: L2 weight decay scaling 0.004, dropout level 0.5, convolution kernel size 1×10, kernel stride 4 and maxpool kernel size 2. Categorical cross entropy is used as the optimisation function with ADAM gradient descent algorithm [107] and ReLU as the activation function on all hidden layers.

4.4.3 Robot Navigation

The experiment begins with the robot facing a scene containing objects which are detected to generate on-screen SSVEP stimuli pixel regions as previously outlined. The subject (teleoperation user) fixates on one particular object from which robot navigation is performed using the high level mobility functions of the NAO humanoid robot platform (Figure 4.1), based on the decoding of the corresponding SSVEP signals by the pre-trained SCU CNN model (Section 4.4.2).

Once these BCI signals are classified as a selected scene object by the subject (user), we then calculate the required robot motion trajectory. As we cannot acquire depth information directly from the monocular camera on the robot, we acquire the distance and the angle of view of the chosen object following the photogrammetric approach of [114]. As such, the distance of the object Z can be calculated as:

$$Z = f' \frac{Y}{y}, \tag{4.1}$$

where Z is the distance in metres, f' is the focal length (pixels), Y is the object height (metres) and y is the object height in the image (pixels). Focal length f' is obtained as:

$$f' = H\frac{f}{h}, \tag{4.2}$$

where H is height of the image (pixels), f is the focal length in metres and h is the camera internal digital sensor height in metres.

The angle of view (AoV) of the object from the camera in radians based on the horizontal position x of the image in pixel can be calculated as follows:

$$AoV = tan^{-1}\frac{x}{Z}. \tag{4.3}$$

When the robot navigates within a given distance and angle trajectory of the subject-selected object, the BCI on-screen interface display alternates to the navigational arrow display (left, right, backwards – Figure 4.1) using the specific SSVEP frequencies of 10, 12 and 15 Hz. These frequencies intend to facilitate robot motion at 90 degree turns left/right or a 180 degree about turn. Subjects similarly attend to one of these SSVEP stimuli which, once decoded by the SCU CNN model, facilitate general robot motion in the environment until further scene objects are detected within the scene traversal. A flow diagram operation of the real-time experimental teleoperation of the NAO robot through the environment in this alternating object-stimuli and navigational-stimuli manner is presented in Figure 4.4.

The experimental navigation plan used during the real-time experiments presented in this study is shown in Figure 4.5. Under these conditions, we repeat the experimental episode five times per subject to demonstrate the repeatability and robustness of our approach.

4.4.4 Offline Dataset

The pre-recorded natural variable stimuli obtained in Section 4.3 is used to gather a larger dataset to investigate the generalisation of such variable stimuli working across subjects. We

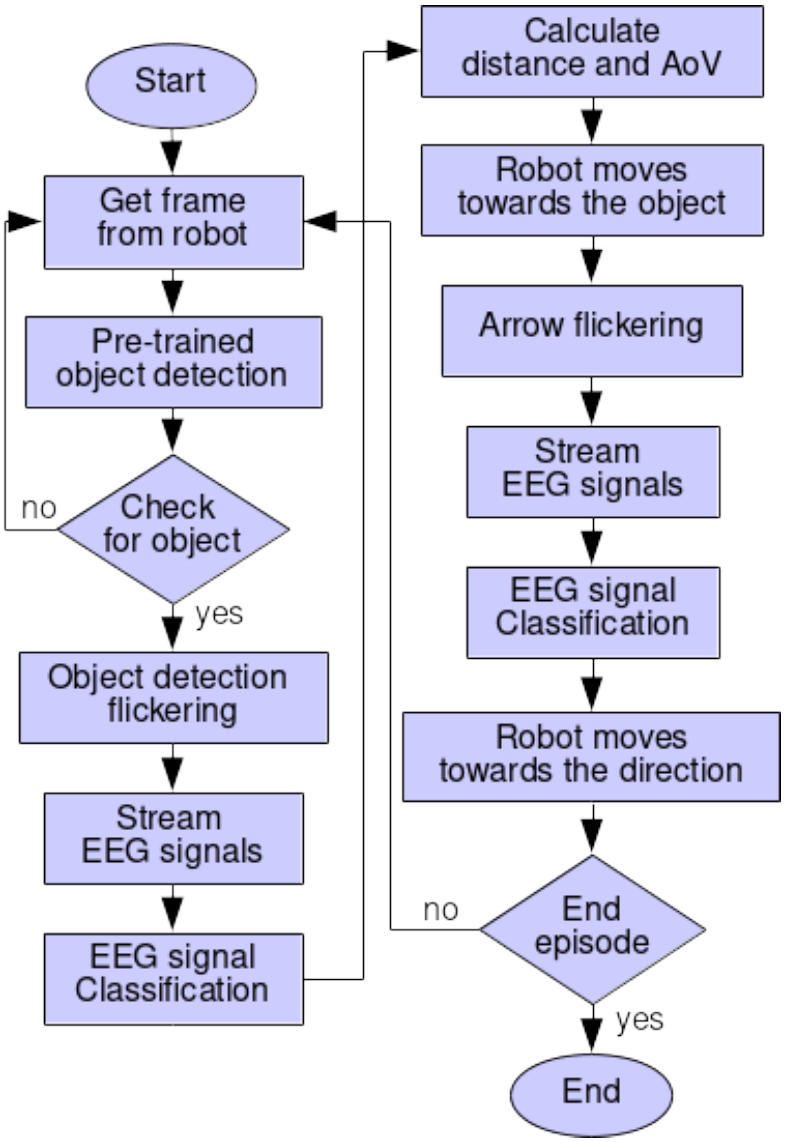


Figure 4.4: Flowchart of real-time robot navigation.

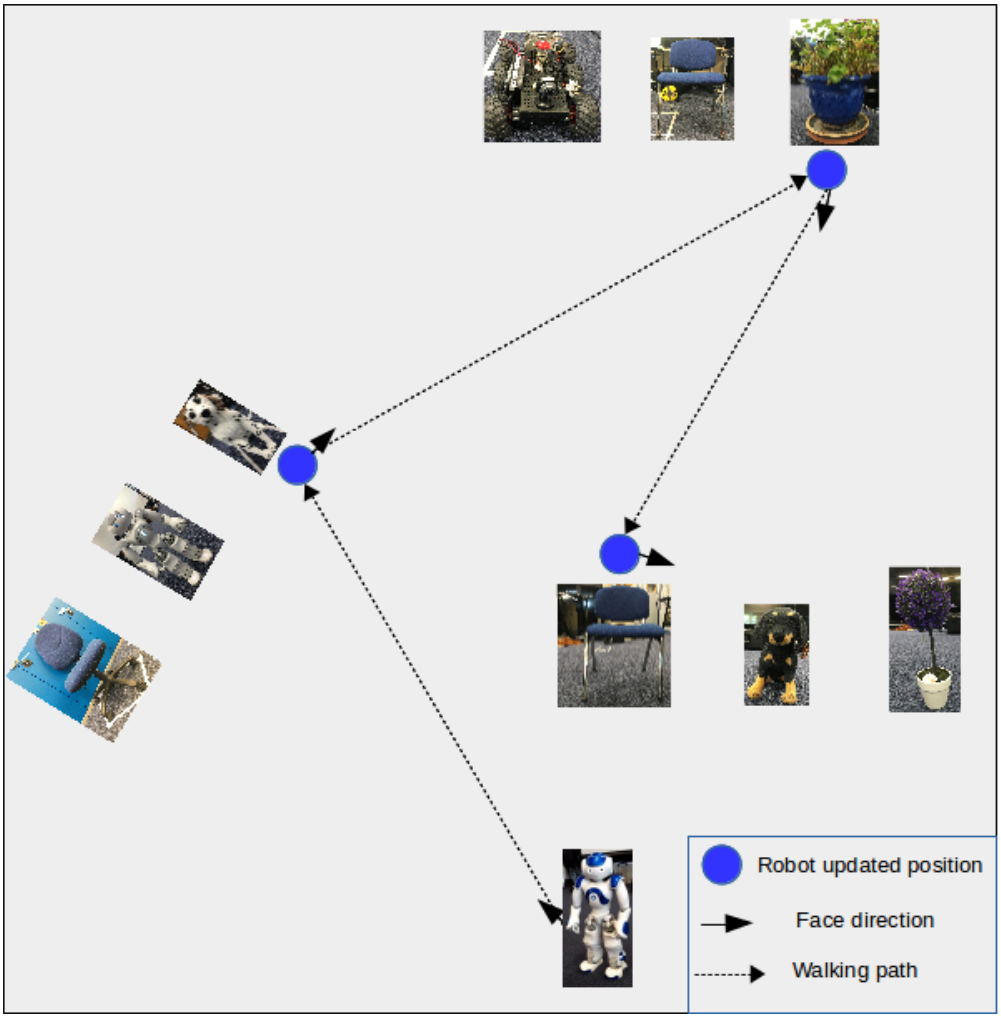


Figure 4.5: Navigation plan for real-time experimentation.

collect the offline dry-EEG dataset from nine subjects (S01 to S09) using the dry-EEG headset (six males and six females) aged between 25 and 40 years old from an offline variable SSVEP stimuli created based on pre-recorded video sequences from a humanoid robot, as detailed in the previous section. We collect 60 samples for each of three classes per each subject.

4.5 Results and Discussion

In this section, we present the results from the calibration stage and the real-time experiment classification using the metrics of classification accuracy and Information Transfer Rate (ITR) in bits per minute (bpm). We also present the results from the offline dataset to re-evaluate the ability of such architectures to classify dry-EEG signals collected via variable SSVEP stimuli. This is able to evaluate our goal of eliminating the rigid EEG-based experimentation, such as the use of traditional wet-EEG system and the non-variable constrain stimuli but yet able to maintain the quality of the bio-signals.

4.5.1 Calibration Statistical Performance

The result for the classification accuracy and the ITR of the calibration experiment are presented in Table 4.1. ITR is the speed of BCI in terms of bit rate transfer, which is the amount of the information throughput by a system per minute [194].

Subject	**S01**	**S02**	**S03**
Accuracy	0.96±0.02	0.80±0.09	0.75±0.12
ITR (bpm)	23.90±0.72	12.16±0.95	9.61±0.89

Table 4.1: Mean accuracy and ITR with standard deviation for calibration classification over 10-fold cross validation.

The ITR is calculated as in [227]:

$$ITR = \frac{B}{T}, \tag{4.4}$$

where T is the time taken to classify a trial in minutes and B is the bits per trial:

$$B = log_2(N) + Plog_2(P) + (1-P)log_2(\frac{1-P}{N-1}), \tag{4.5}$$

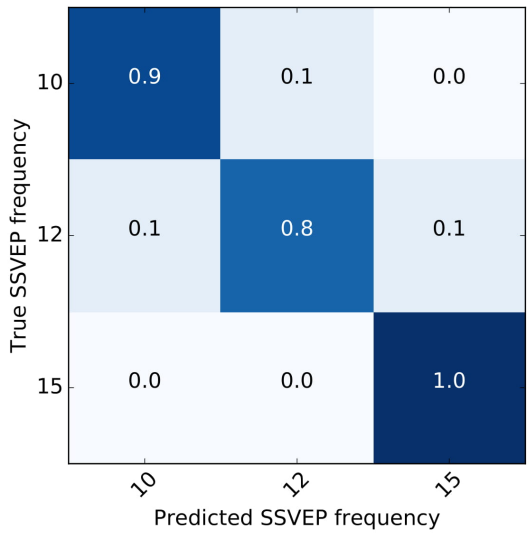
where N is the number of possible selections (N = 3) and P is the correct selection accuracy. ITR is a suitable BCI performance metric, as ITR indicates how fast the accurate information can be transferred. As can be seen in Equation 4.5, the number of bits per trial, B is dependent upon high accuracy of P.

For the calibration experiment, the time taken is based on the total flickering time per trial (3 seconds) plus the average of time the classifier takes to train and classify a trial. The data collected during the calibration experimental phase is used to train the model for real-time experimentation. However, in order to demonstrate statistical performance of our SCU CNN architecture on this task, we present mean accuracy over 10-fold cross validation per subject (Table 4.1). This is used as the P value to calculate the value for B within the overall ITR calculation (Equation 4.5).

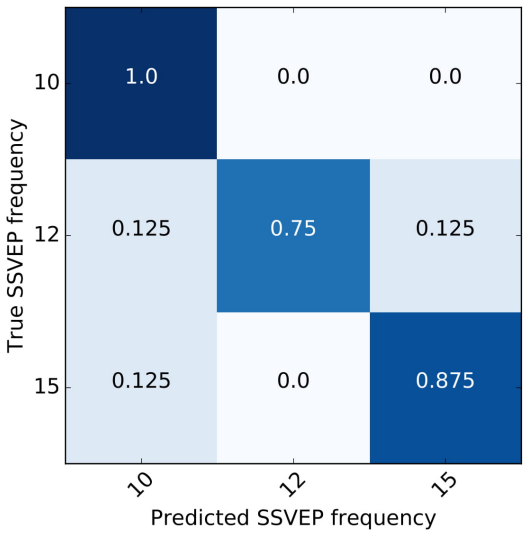
4.5.2 On-line Real-time Performance

The results of the on-line experimental phase are presented in Table 4.2 where we can see the results are depending on the performance of signals decoding during calibration stage. This is because the training weights for on-line real-time classification are from training the signals collected during calibration stage. Overall, the results demonstrate extremely high accuracy for all of the subjects tested.

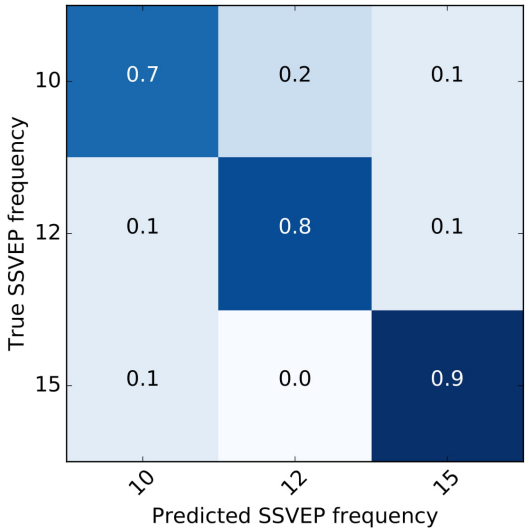
Our results demonstrate a strong statistical performance, with a mean accuracy of 0.85 across all subjects. This is comparable to [258] which obtained 0.88 accuracy, despite our work using a variable SSVEP stimuli. As ITR represents the speed of the real-time information transfer from stimuli to motion command generation, the time taken is measured from the beginning of a stimuli flashing until getting a prediction. We can thus improve the ITR further via reducing the flickering time during the real-time experiment.



(a) Subject S01.



(b) Subject S02.



(c) Subject S03.

Figure 4.6: Confusion matrices for the classification of real-time EEG signals during the robot navigation (maximal result being *accuracy* = 1.0 in the matrix diagonals).

Subject	S01	S02	S03
Experiment 1	1.00	1.00	1.00
Experiment 2	1.00	0.83	0.50
Experiment 3	0.83	0.83	0.83
Experiment 4	0.83	1.00	1.00
Experiment 5	0.83	0.67	0.67
Mean Accuracy	0.90±0.08	0.87±0.12	0.80±0.19
Mean ITR (bpm)	16.8±0.10	15.6±0.12	13.2±0.16

Table 4.2: Accuracy for each experiment and mean accuracy and ITR with standard deviation for real-time classification.

Figure 4.6 represents per-class confusion matrices for the real-time classification and highlights overall good accuracy across all classes for all three subjects (users), although the middle class (12 Hz) is more difficult to classify than the rest of the classes.

Figure 4.7 illustrates the real-time experimental environment such as the view from the robot and the robot approaching an object. The angle of direction from the robot to the selected objects can vary from one experiment to another, because the calculation of distance and direction is based on the bounding box from the object detection and the angle of view of an object on the plane. The detected bounding box for the scene object can vary and the angle of view of an object can change with the slightest movement of either the robot or the robot head (where the camera is located).

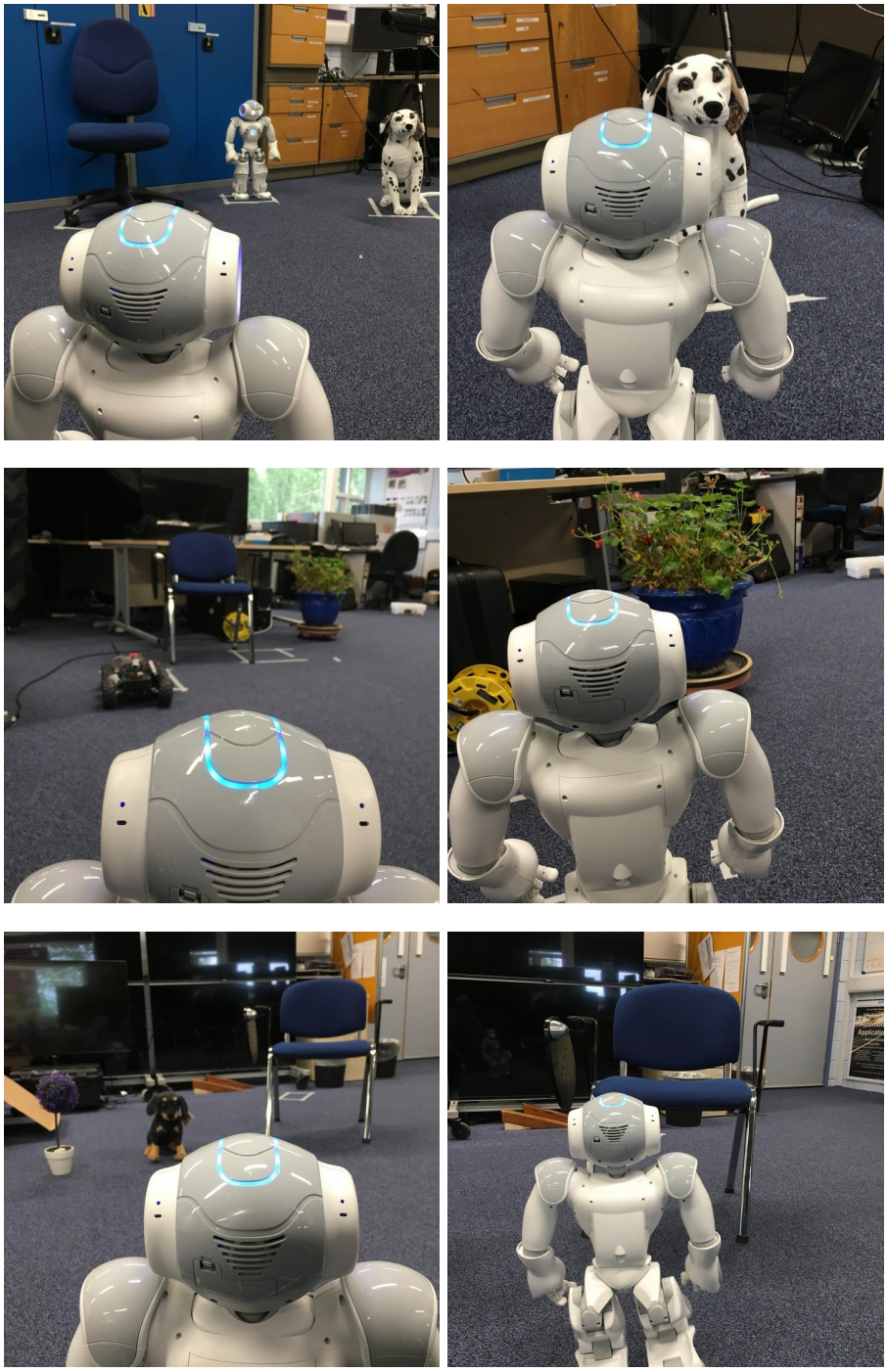


Figure 4.7: Sample of humanoid robot navigation during a real-time experiment.

4.5.3 Offline Dataset Performance

In this section, we present the experimental evaluation of our SCU approach (as in Chapter 3) to accurately classify dry-EEG signals collected via a variable SSVEP stimuli. We evaluate our SCU approach against RNN models including a vanilla RNN, LSTM and GRU. All the presented results are produced via 10-fold cross validation, with models which were initially optimised using hyperparameters chosen via a grid-search over a validation set.

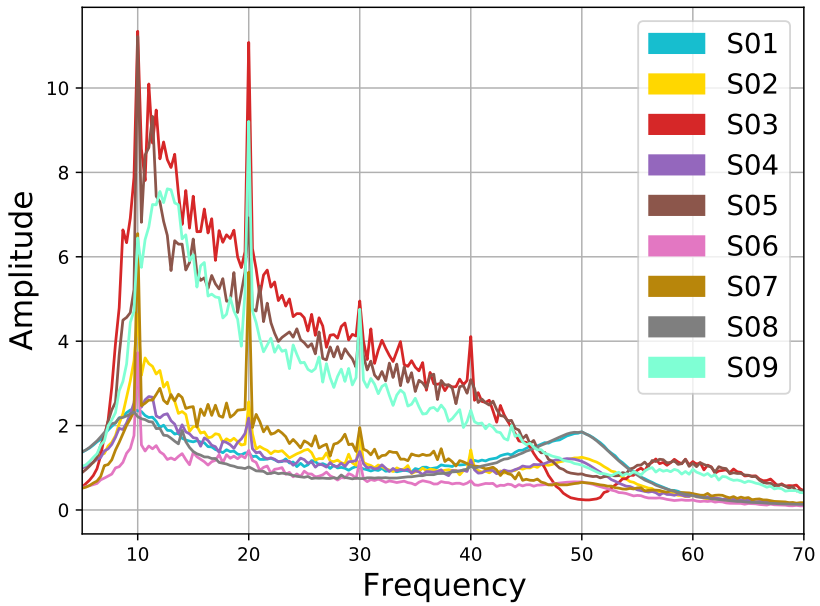
To ensure the collected signals contain clearly identifiable target frequencies, in Figure 4.8, we visualise the collected dataset per subject and per class via an FFT. FFT is used to decompose the EEG signals into the frequency domain to demonstrate the characteristic SSVEP peaks at the target frequency and associated harmonics [194].

SSVEP Classification for a Single Subject

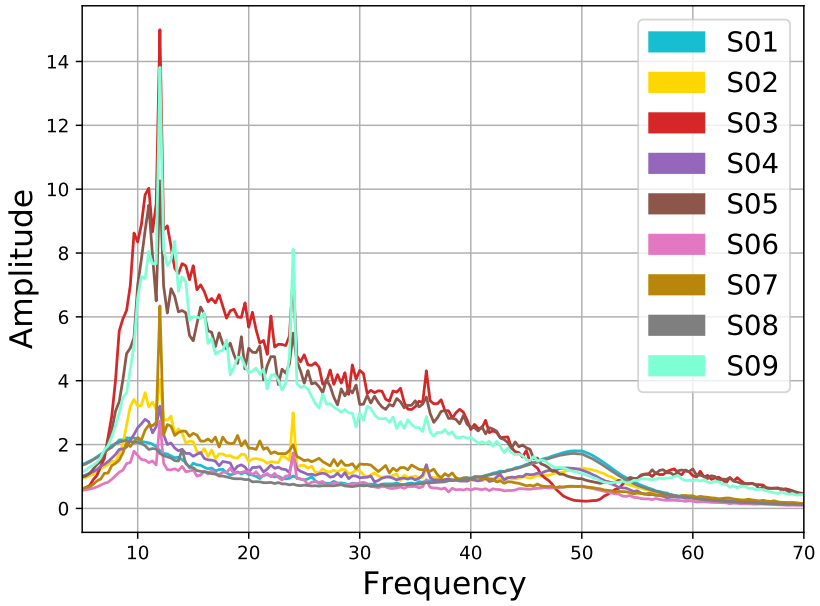
For this experiment, we attempt to classify SSVEP signals using data from a single subject – this means that a new model is trained for each of the nine subjects which were collected offline using our proposed novel variable stimuli. The results from this experiment are presented in Table 4.3. The results highlight, that even with more complex stimuli than the one used in previous chapter (Chapter 3), our SCU approach still outperforms the RNN models by a large margin.

SSVEP Classification Across Subjects

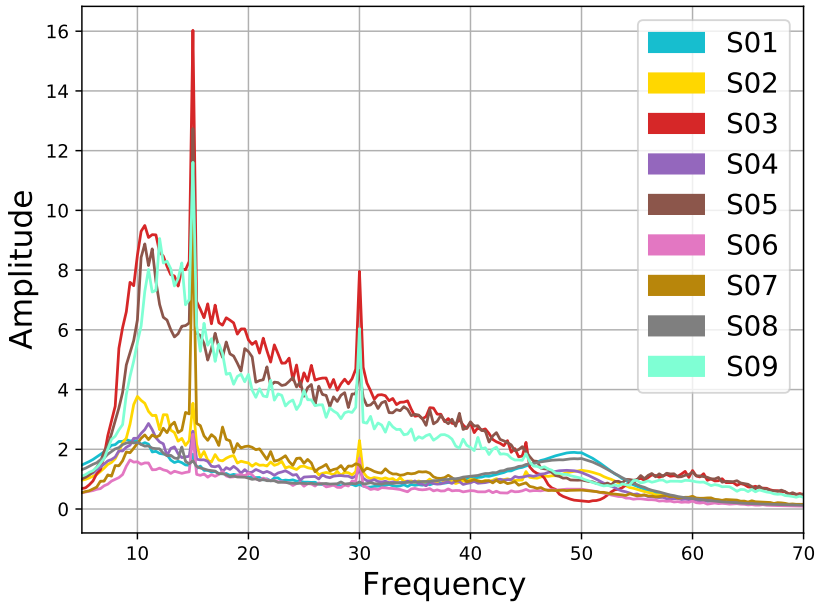
Having a single model trained on EEG signals taken from multiple subjects is known to be challenging [56], potentially due to biological differences between subjects and the variability of the EEG recording process. To assess the ability of our SCU model to classify a dataset comprising data from all of nine subjects, we classify all the signals from the nine subjects together instead of performing individual classification. The results presented in Table 4.4 show that our SCU model is able to generalise across all nine subjects when performing classification across subjects.



(a) FFT for class frequency equal to 10Hz.



(b) FFT for class frequency equal to 12Hz.



(c) FFT for class frequency equal to 15Hz.

Figure 4.8: FFT when taking across samples on all nine subjects. The plots show the peak frequency and the harmonics of all three frequencies.

Method	SCU	RNN	LSTM	GRU
S01	**0.648 ± 0.075**	0.333 ± 0.033	0.322 ± 0.060	0.344 ± 0.056
S02	**0.784 ± 0.078**	0.339 ± 0.021	0.358 ± 0.059	0.444 ± 0.094
S03	**0.816 ± 0.053**	0.347 ± 0.028	0.375 ± 0.043	0.361 ± 0.066
S04	**0.719 ± 0.050**	0.331 ± 0.052	0.314 ± 0.033	0.333 ± 0.071
S05	**0.807 ± 0.032**	0.331 ± 0.029	0.358 ± 0.047	0.464 ± 0.061
S06	**0.873 ± 0.041**	0.317 ± 0.061	0.331 ± 0.050	0.461 ± 0.076
S07	**0.878 ± 0.045**	0.319 ± 0.043	0.375 ± 0.055	0.500 ± 0.115
S08	**0.713 ± 0.122**	0.319 ± 0.031	0.286 ± 0.066	0.313 ± 0.069
S09	**0.858 ± 0.050**	0.342 ± 0.072	0.356 ± 0.035	0.450 ± 0.081
Mean	**0.789 ± 0.080**	0.331 ± 0.011	0.342 ± 0.030	0.408 ± 0.069

Table 4.3: Mean accuracy with standard deviation over 10-fold cross validation for nine dry-EEG subjects.

Method	Accuracy
SCU	**0.650 ± 0.029**
RNN	0.333 ± 0.031
LSTM	0.382 ± 0.062
GRU	0.382 ± 0.056

Table 4.4: Test Accuracy across subjects

SSVEP Classification for Unseen Subject

The pattern of electrical voltage produced by the brain often differs from one subject to the next, resulting in difficulty for a model to correctly classify SSVEP signals from a subject whose signals is absent from a priori model training. Having a model that able to generalise the SSVEP classification of dry-EEG signals on unseen subject is a strong desirable quality. We evaluate the capability of our SCU approach to generalise the trained features to an unseen subject (subject whose data is absent *a priori* model training). We utilise the Leave-one-out methods to train the network on eights subjects and test on an unseen subject. From the previous result in Section 4.5.3 and Section 4.5.3, we only consider the highest performing classification approach (our proposed SCU) for the unseen subject evaluation.

Method	1SCU
Unseen S01	0.471 ± 0.046
Unseen S02	0.719 ± 0.014
Unseen S03	0.679 ± 0.011
Unseen S04	0.675 ± 0.016
Unseen S05	0.743 ± 0.009
Unseen S06	0.703 ± 0.101
Unseen S07	0.885 ± 0.063
Unseen S08	0.425 ± 0.029
Unseen S09	0.806 ± 0.023
Mean	0.678 ± 0.147

Table 4.5: Mean accuracy with standard deviation when classifying SSVEP for unseen subject by leave-one-subject-out.

4.6 Conclusion

In this chapter, we have presented a number of novel contributions spanning the use of variable SSVEP stimuli (pattern, size, shape) as an enabler of future telepresence BCI applications in a real-world natural environment. In comparison to the prior work [258, 137, 38, 210, 227], these variable stimuli vary in terms of pixel pattern, size and on-screen position in-conjunction with the changing nature of the environment the robot is navigating through. We integrate recent advances in the use of deep CNN architectures for both scene object detection and dry-EEG bio-signal decoding. Within this context, we developed a novel SSVEP interface to flicker the on-screen frequency of naturally occurring objects detected within the scene, as seen from the on-board camera of a teleoperated robot and decode these dry-EEG brain-based bio-signals based on the frequency of the visual fixation detected to navigate the robot within the scene.

Uniquely, we train and utilize a common CNN model (SCU, Figure 4.3) for use with SSVEP stimuli that vary in size, on-screen position and internal (pixel pattern) throughout the duration of the experiment, significantly advancing such decoding generality against prior work in the field [210, 227]. Our evaluation is presented in terms of accuracy and ITR, both on the *a priori* experimental training set used for the off-line training phase (via cross-validation) and the on-line real-time teleoperated navigation of a humanoid robot through a natural indoor environment. The introduction of these highly novel and variable BCI SSVEP stimuli, based on

scene object occurrence, demonstrates adaptable BCI-driven robot teleoperation within a natural environment (without scene markers and alike). Strong statistical classification performance is observed, comparable to and often exceeding those reported in the general BCI literature [258], despite the introduction of the serious challenges associated with variable SSVEP stimuli.

Whilst the proposed stimuli and associated experimentation utilised in this chapter is novel, the results presented in this chapter are based on a controlled environment where a preprepared floor plan is used for the robot navigation. This is due to the limitation of the on-board camera on our humanoid robot that only provides a 2D image. On top of that, the experiment has been conducted within an indoor environment with the use of a default condition where the humanoid robot will always move towards the ground truth label. However, any incorrect predictions have been incorporated in the reported accuracy score.

Potential future work will look to improve generalisation performance over additional test subjects, increase both scene complexity and teleoperated duration as well as considering aspects of robot interaction within the environment. Perhaps by incorporating a depth camera, the experiment can be more flexible with the depth information used to calculate the distance and the direction instead of the pre-determined height of the object. Perhaps it would also be interesting to investigate if it is possible to navigate a mobile robot in an outdoor environment using the proposed novel variable stimuli.

Chapter 5

Generating Brain Signals

Substantial progress has been made in the area of BCI using modern machine learning techniques to decode and interpret brain signals. While EEG has provided a non-invasive method of interfacing with a human brain, the acquired data is often heavily subject and session dependent. This makes seamless incorporation of such data into real-world applications intractable as the subject and session data variance can lead to long and tedious calibration requirements and cross-subject generalisation issues. This implies that many downstream applications, including Steady State Visual Evoked Potential (SSVEP) based classification systems, can suffer from a shortage of reliable data.

In this chapter, taking into consideration the problem of data collection during EEG-based BCI experiments such as being expensive and time-consuming, we tackle the problem of the shortage of reliable data by alternatively generating meaningful and realistic synthetic data, which can therefore be of significant value in circumventing this problem. In this chapter, we explore the use of modern neural-based generative models trained on a limited quantity of EEG signals collected from different subjects to generate supplementary synthetic EEG signals, subsequently utilised to train an SSVEP classifier.

The work presented in this chapter has been published in the following peer-reviewed publication:

Nik Khadijah Nik Aznan, Amir Atapour-Abarghouei, Stephen Bonner, Jason D Connolly, Noura Al Moubayed, and Toby P Breckon. Simulating Brain Signals: Creating Synthetic EEG Data via Neural-based Generative Models for Improved SSVEP Classification. In International Joint Conference on Neural Networks*, pages 1–8. IEEE, 2019*

5.1 Introduction

BCI is a system that translates such acquired signals to provide a communication and control medium between the human brain and external devices. BCI has received significant attention within the research community for decades [142]. However, not many BCI applications are tractable in daily use in real-world scenarios, especially for important medical applications, such as assisting patients with locked-in syndrome, as there are numerous shortcomings and limitations within the current state of the art that lead to the low reliability and usability of BCI [140].

Deep neural networks have recently been used to improve the classification of various aspects of EEG signals [115, 226]. While it is essential to have access to large quantities of data for training such methods, collecting high quality EEG signals has proven difficult [253, 145]. This can be for a variety of reasons including the requirement for careful per-subject and per-session calibration. This makes EEG BCI experiments time-consuming, expensive and difficult to operate within the usually short amount of time experimental subjects can perform EEG experiments [53]. In addition to these issues, there are commonly-known limitations with EEG signals in general, which can severely hinder the applicability of a system dependent on such data [253]. For instance, EEG signals are known to be highly subject and session variant, which leads to a long calibration process for every individual experiment [140, 142, 10]. This further impacts any machine learning based models which are trained upon this data, as they often demonstrate poor generalisation performance across different subjects or experiential data collecting sessions [140, 226].

In machine learning, generative models have long been used to generate entirely new and realistic data points which match the distribution of a given target dataset [205]. Recent work on neural-based models such as Generative Adversarial Networks (GAN) and Variational Auto-Encoders (VAE) have demonstrated that these are highly capable at capturing key elements from a diverse range of datasets to generate realistic samples [82]. Increasingly, there is evidence that

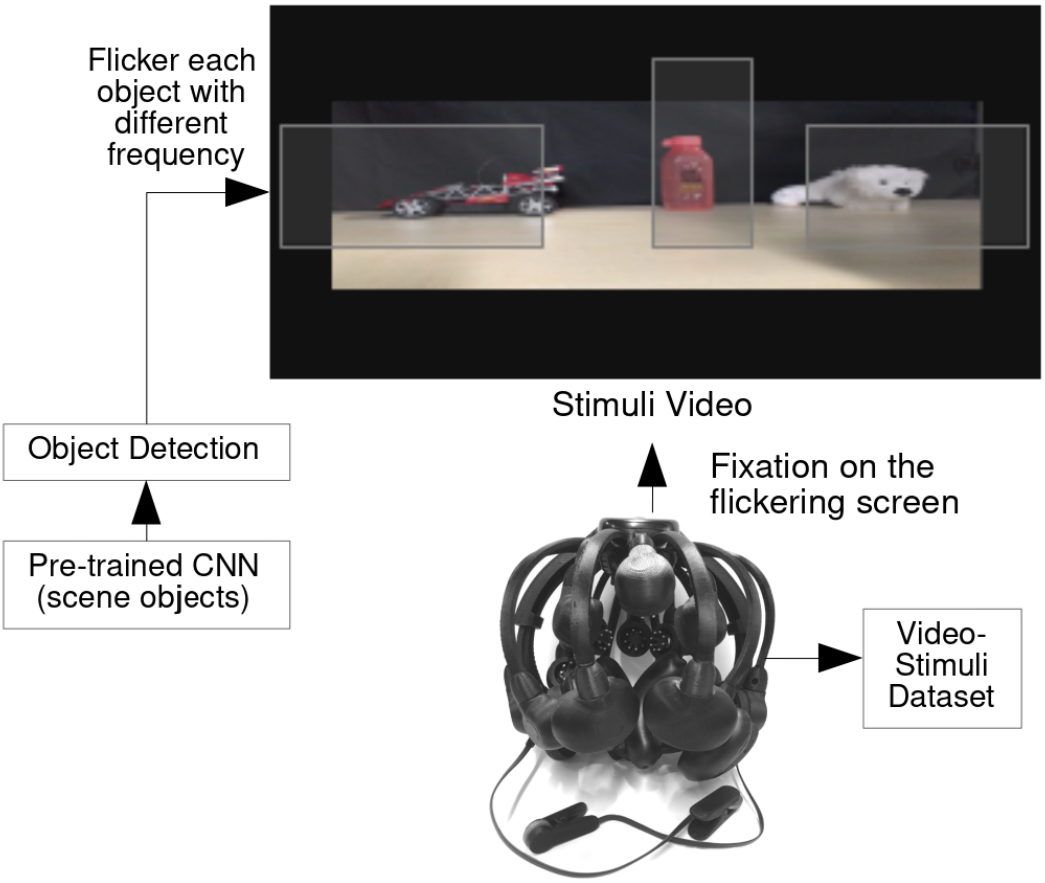


Figure 5.1: Details of the collection process for an SSVEP EEG dataset (*Video-Stimuli*) using a video based stimuli.

using synthetic data, taken from a generative model, can be used as a form of data augmentation to help improve the performance of any down-stream data classification task [72].

In this chapter, we detail the generation of new synthetic EEG signals using a selection of customised neural-based generative models and explore applications of such data including using it to boost classification accuracy on real datasets. Uniquely, we specifically focus on dry-EEG signals containing Steady State Visual Evoked Potential (SSVEP) signals. Dry-EEG requires no conductive gel which improves the usability of within a BCI context, eliminating the major limitation of the wet EEG system [62, 139, 160]. However, dry-EEG results in high impedance

values that causes more noise and artefacts in the data, creating more challenging signal decoding and classification [62].

To the best of our knowledge, this is the first work in the literature to explore the use of neural-based generative models to create dry-EEG signals containing SSVEP information. In summary, we make the following major contributions in this work:

- The generation of synthetic dry-EEG signals containing SSVEP signals using a variety of unsupervised models.
- A demonstration that using generated data can improve the classification of real-world EEG signals, taken from multiple subjects and recorded under various conditions and sessions.
- An exploration of both classifier pre-training and dataset augmentation as use cases for the generated data.
- Further demonstration that using synthetic EEG signals can increase the convergence rate of classification models, thus resulting in the observation that smaller quantities of real-world training data is required.

We perform extensive experiential evaluations to validate our claims and to aid reproducibility, we release our Python-based (PyTorch) implementation of all the generative models, along with sample input data here - https://github.com/nikk-nikaznan/SSVEP-Neural-Generative-Models.

5.2 Related Work

We consider the background information relevant to this work within two distinct areas:- recent advances made in generative models (Section 5.2.1) and existing work on using such models to generate meaningful and coherent synthetic EEG signals (Section 5.2.2).

5.2.1 Neural-Based Generative Models

Generative models have been proven very powerful within the context of unsupervised learning where the model learns a hidden structure of the data from its distribution to generate new data samples within the same distribution [99]. This generated dataset often contains enough variation to support the down-stream training of a secondary model [145].

Generative Adversarial Networks (GAN) [82] are capable of producing semantically sound artificial samples by inducing a competition between a generator (G), which attempts to capture the distribution, and a discriminator (D), which assesses the generator output and penalizes unrealistic samples. Both networks are trained simultaneously to achieve an equilibrium.

Training a GAN is known to be challenging with pervasive instability issues [8]. One such issue stems from the discriminator rapidly reaching optimality and effortlessly distinguishing between the fake samples output by the generator and samples from the real distribution. This will lead to a lack of meaningful gradients for training, effectively ceasing any progress towards the equilibrium.

The Wasserstein GAN (WGAN) is consequently proposed in [9] to rectify some of the issues associated with training a GAN. The Wasserstein-1 metric is used to measure the distance between the real and model distributions. Also known as the Earth Mover's distance, ($EM(p, q)$), this metric is the minimum cost of moving distribution elements (earth mass) to transform a distribution q to distribution p (cost = mass $\times$ transport distance).

The Wasserstein GAN [9] has an aptly named *critic* (C) instead of a discriminator since this network is no longer a classifier. Using the EM distance, the critic will not only determine whether a sample is fake or real as a discrete binary decision, but how real or how fake the generated sample is as a continuous regressive output. Under the right training circumstances, the critic will eventually converge to a linear function with ever-present meaningful gradients and cannot saturate. The loss function in the Wasserstein GAN is created via the Kantorovich-Rubinstein duality [9]:

$$\min_{G} \max_{C \in \mathcal{F}} \mathop{\mathbb{E}}_{x \sim \mathbb{P}_r} [C(x)] - \mathop{\mathbb{E}}_{\tilde{x} \sim \mathbb{P}_g} [C(\tilde{x})], \tag{5.1}$$

where $\mathcal{F}$ is the set of 1-Lipschitz functions, $\mathbb{P}_r$ the real distribution, $\mathbb{P}_g$ the model distribution defined by $\tilde{x} = G(z), z \sim p(z)$, and z the random noise. If C is optimal, minimizing the value function with respect to G minimizes $EM(\mathbb{P}_r, \mathbb{P}_g)$.

The Wasserstein GAN does not suffer from vanishing gradients or mode collapse. However, to guarantee continuity, a Lipschitz constraint must be enforced, which is achieved in [9] by *clamping* the weights. This results in the creation of a new hyper-parameter, which needs to be carefully tuned to the distribution.

A gradient norm penalty with respect to the critic input is consequently proposed in [85] to replace clamping. Since a differentiable function is 1-Lipschitz if and only if its gradient norm is no more than 1 everywhere, [85] limits the critic gradient norm by penalizing the function on the gradient norm for samples $\hat{x} \sim \mathbb{P}_{\hat{x}}$, where $\hat{x} = \epsilon x + (1 - \epsilon)\tilde{x},\ 0 < \epsilon < 1$. This penalty term which is added to the function in Eqn. 5.1 is therefore as follows [85]:

$$\underset{\hat{x} \sim \mathbb{P}_{\hat{x}}}{\mathbb{E}} [(||\nabla_{\hat{x}} C(\hat{x})||_2 - 1)^2], \tag{5.2}$$

where $\mathbb{P}_{\hat{x}}$ is implicitly defined to sample uniformly along straight lines between pairs of points sampled from the real data distribution, $\mathbb{P}_r$, and $\mathbb{P}_g$ is the model distribution defined by $\tilde{x} = G(z), z \sim p(z)$ [85].

Auto-encoders have long been used as a method of creating a low-dimensional representation z of data using an encoder model, which can be used to reconstruct the original data with minimal errors via a decoder model [92]. However, traditionally they cannot be explicitly used to generate new data samples based on the learned data distribution. Variational Auto-encoders (VAE) utilise ideas from Bayesian inference to produce a more expressive data representation, whilst also having the ability to generate new data samples [108, 198]. Unlike non-probabilistic auto-encoders [92], a VAE does not learn a fixed value for each element in z but instead each element is sampled from a probability distribution before being passed to the decoder model. This has been shown to produce a more semantically meaningful representation, where individual dimensions in the hidden space can correspond to tangible elements in the dataset, such as facial expression in a dataset of human faces [119].

As the decoder model of the VAE is trained to take a sample from a Gaussian distribution and produce a realistic output, it can be used to produced new data by simply sampling points in the distribution and reconstructing them.

5.2.2 Literature Review

Within the existing body of work in the literature, there are instances of generative models used to create synthetic EEG signals. For instance, [89] proposes using the improved WGAN with Gradient Penalty (GP) for a single channel of the EEG signals for motor imagery task. The generator consists of one linear layer, six convolutional layers where each layer consists of an up-sampling operation, two convolutions and one fully-connected layer. The authors evaluate the performance of four models, three different up-sampling methods (nearest-neighbour, linear interpolation and cubic interpolation with convolutional down-sampling) and down-sampling via average pooling with the original WGAN-GP. All the models are then assessed using four evaluation metrics; inception score, Frechet inception distance, Euclidean distance and sliced Wasserstein distance. All four models are demonstrated to outperform WGAN-GP.

In [53], the authors propose deep EEG super-resolution using a GAN. The model is applied to a small number of EEG channel data to interpolate other channel signals using motor imagery dataset from [159]. The super-resolution data (SR) is generated via WGAN with convolutional filters using the low-resolution data (LR) from the dataset by down-sampling the EEG channels by scale factors of two and four for two different experiments. The channels removed from the down-sampling processes are used as high-resolution data (HR) to compete against SR in the discriminator. They evaluate the performance of the SR by performing classification and comparing the accuracy with the classification performed using HR. The authors conclude that SR is capable of producing high spatial resolution EEG signals from low resolution signals.

Instead of generating EEG signals, [180] generates previously seen images while having brain signals recorded by EEG. Six subjects are shown 2000 images with 40 classes per subject. The EEG signals are pre-processed by hardware notch filter between 49 and 51 Hz and bandpass filter between 14 and 70 Hz. Using LSTM RNN, the temporal feature representations are encoded, which are subsequently used as condition vectors employed along with random noise vectors to generate new images. They obtain a test accuracy of 83.9% when evaluating the LSTM model for feature representation.

In [253], synthetic EEG signals are generated by inverting the artificial Time-frequency Representation (TFR) obtained from conditional Deep Convolutional GAN (cDCGAN). The Wavelet Transform is used to obtain the TFR of the signals to be used by the cDCGAN to generate the artificial TFR of the EEG signals. The BCI competition II dataset III [209] is used as the EEG signals. They evaluate the efficacy of the synthetic data by comparing the classification accuracy of the model trained on real data, synthetic data and a mixture of real and synthetic data with different ratios. Using additional synthetic data, the accuracy improved up to 3%.

In [145] and [191], a GAN is used to generate synthetic emotion recognition EEG signals. The authors in [145] generate EEG in the form of differential entropy from noise distribution using a conditional WGAN with two emotion recognition datasets. They evaluate the performance of the generated data by combining the synthetic and real-world data to train a classification model compared against a model solely trained on real-world data. The addition of the synthetic data leads to an improvement in the accuracy of up to 20% for different datasets.

The approach in [191] combines classification and generative networks in a model using an Auxiliary Classifier GAN (AC-GAN). Part of the approach is encoding the data from SEED [262] and DEAP [109] datasets into images using a Markov Transition Field (MTF) before passing them into the AC-GAN to generate new data samples. Every sample comes with a corresponding class label. A Tiled CNN is employed to classify the MTF images to either fake or real and with the class label. They improved the classification by less than 1% in the SEED dataset as compared with previous work.

In this paper, we generate synthetic EEG signals from SSVEP tasks using dry-EEG. While earlier work [253, 191] has been able to generate features within a secondary domain, from which EEG signals are subsequently reconstructed, we are the first to introduce the concept of generating meaningful EEG signals directly in signal space via end-to-end training instead of first transforming the signals into different domains for SSVEP classification. This improves overall efficiency, removes any need for additional signal processing, and prevents potential instability errors introduced during transformations.

5.3 Proposed Generative Networks

As one of the objectives of this work is to create synthetic EEG signals, which can be used to improve the training of a downstream classifier [15], we explore the use of a variety of neural-based generative models (as described in Section 5.2). In this section, a brief description of the details of the models used in this work is provided.

5.3.1 Deep Convolutional Generative Adversarial Network

As part of this work, we use the generative model proposed in [192]. Random noise vectors z are sampled from a Gaussian distribution and used as the generator input. At every iteration, the generator outputs fake data samples ($\tilde{x} = G(z)$), which are then passed to the discriminator along with randomly selected real data samples x, classifying them as either fake or real, the gradients from which are employed to train the generator, leading to higher quality outputs at every step. An overview of our generative adversarial model is seen in Figure 5.2.

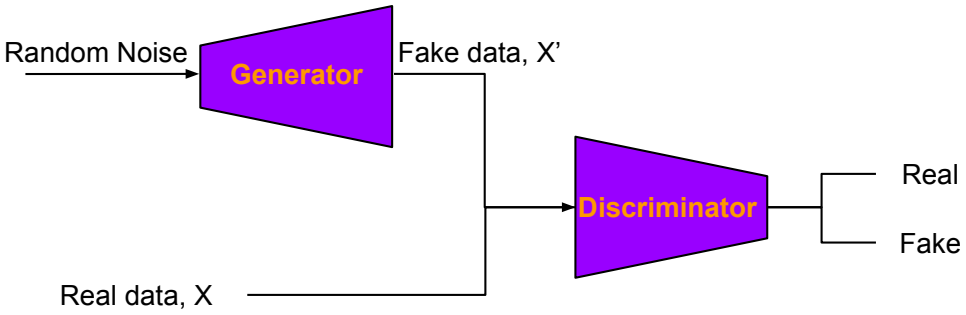


Figure 5.2: An overview of the generative adversarial network.

Loss Function The loss function is based on the competition between the generator and the discriminator following the minimax objective [82]:

$$L = \min_{G} \max_{D} \mathop{\mathbb{E}}_{x \sim \mathbb{P}_r} [log(D(x))] + \mathop{\mathbb{E}}_{\tilde{x} \sim \mathbb{P}_g} [log(1 - D(\tilde{x}))], \tag{5.3}$$

where $\mathbb{P}_r$ is the data distribution, $\mathbb{P}_g$ the model distribution defined by $\tilde{x} = G(z), z \sim p(z)$, and z the random noise vector used as the input to the generator. An overview of the training pipeline is seen in Figure 5.2.

Implementation Details The network architecture is based on that of [192], with the exception of the use of one dimensional convolutions since the networks process EEG signal vectors rather than images. Our generator consists of one dense layer and three 1D transpose convolutional layers as in Figure 5.3.

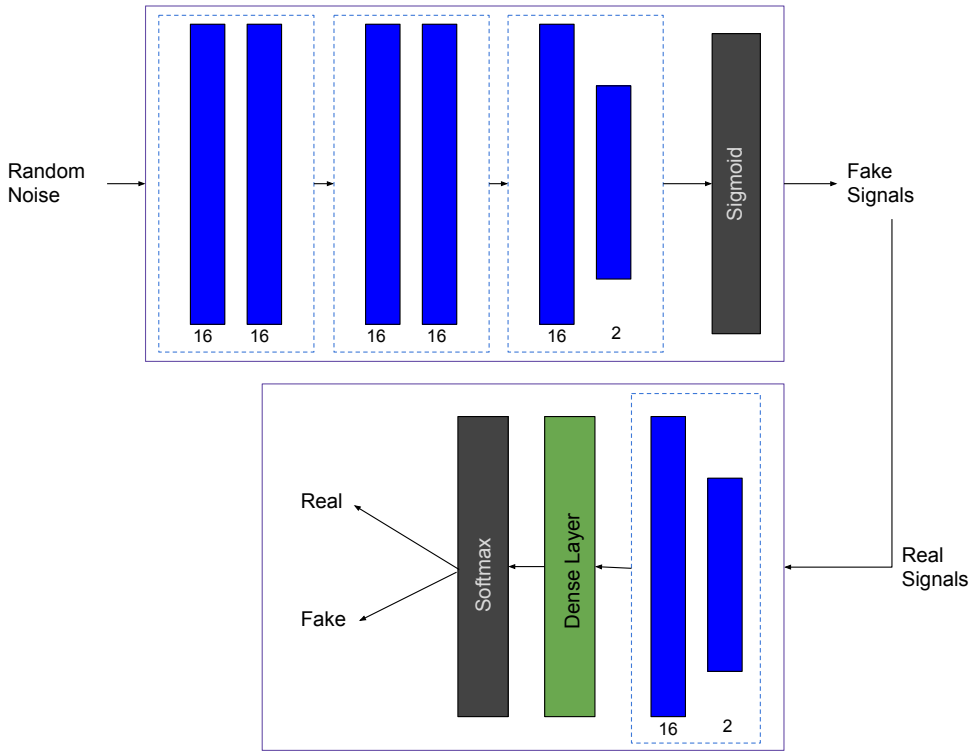


Figure 5.3: Our proposed GAN.

This vector is then used as the input to our light-weight discriminator, consisting of two layers; the first containing a convolutional, batch normalised leaky ReLU activation layer with parameter slope=0.2, followed by max pooling, and the second using a fully-connected layer followed by a leaky ReLU ($slope = 0.2$).

Implementation and training is carried out in *PyTorch* [181], with Adam [107] used as the optimization approach (momentum $\beta_1 = 0.5$, $\beta_2 = 0.999$, initial learning rate $\alpha = 0.0001$).

5.3.2 Improved Wasserstein Generative Adversarial Network

Similarly to the training procedure in Section 5.3.1, the Wasserstein GAN is made up of two completing networks, a generator and a critic. The generator receives a random noise vector z as its input, and the critic determines how real or fake the data samples created by the generator are by calculating the distance between the real data distribution and the model distribution (Earth Mover's distance).

Since it is significantly important to keep the critic optimal at all times, we train the critic 25 times per each generator training iteration for the first 100 generator iterations and 5 times per each generator iteration for the rest of the training process.

Loss Function Here, we take advantage of the improved Wasserstein GAN [85] with the following loss function:

$$\begin{aligned} L = \min_G \max_C \mathop{\mathbb{E}}_{\tilde{x}\sim\mathbb{P}_g}[C(\tilde{x})] - \mathop{\mathbb{E}}_{x\sim\mathbb{P}_r}[C(x)] + \\ \lambda \mathop{\mathbb{E}}_{\hat{x}\sim\mathbb{P}_{\hat{x}}}[(||\nabla_{\hat{x}} C(\hat{x})||_2 - 1)^2], \end{aligned} \tag{5.4}$$

where $\mathbb{P}_g$ is the model distribution defined by $\tilde{x} = G(z), z \sim p(z)$, z is random noise, $\mathbb{P}_r$ is the true data distribution, and $\mathbb{P}_{\hat{x}}$ is implicitly defined to sample uniformly along straight lines between pairs of points sampled from $\mathbb{P}_r$ and $\mathbb{P}_g$.

Implementation Details For the sake of consistency, the architecture of the networks used here is similar to that of the networks in Section 5.3.1 where all the architecture layers and hyperparameters are the same except for the loss function. Similarly, all implementation and training is carried out in *PyTorch* [181], with Adam [107] used as the optimization approach (momentum $\beta_1 = 0.1$, $\beta_2 = 0.999$, initial learning rate $\alpha = 0.0001$).

5.3.3 Variational Autoencoder

In addition to the GAN based approaches, we also explore the creation of a Convolutional Variational Auto-encoder (detailed in Section 2.3.3) to generate synthetic EEG signals. Our VAE model uses 1D convolutions to encode features from a given EEG signals sample, which are used to parametrise a Gaussian distribution from which a latent and compressed representation of the input data is sampled. This latent representation is passed to the decoder section of the model which comprises transposed convolutions used to transform the latent representation back into the original EEG signals.

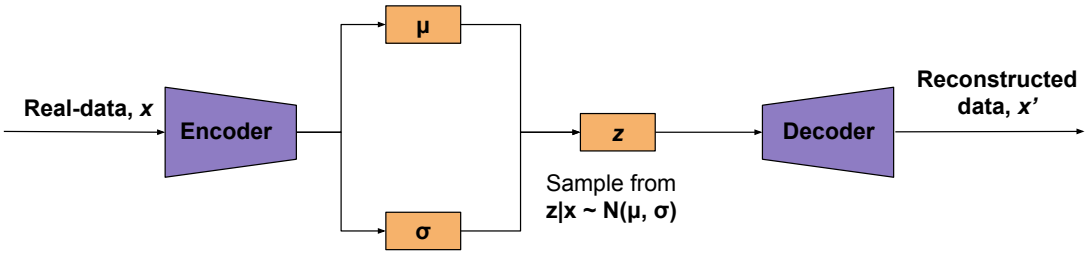


Figure 5.4: An overview of the Variational Auto-encoder.

Once we have trained the VAE model using the objective function detailed in the next section, we are able to use it to generate entirely new data samples, which could have plausibly come from the training data, but which is not conditional on any input to the model.

Loss Function The encoder section of our VAE model is trained to learn two output vectors, μ and σ, which represent the mean and variance of the latent space from which z will be sampled, $z = \mathcal{N}(\mu, \sigma)$. Using the sampled representation z, the decoder section of the VAE is trained to reconstruct the input data x. Consequently, our VAE is trained to infer the intractable distribution $p(z|x)$, that being the likelihood of z given the data $\hat{x}$, using a stand-in tractable distribution $q(z|x)$ in the following manner:

$$L = \mathbb{E}_{q(z|x)}\Big[log(p(x|z))\Big] - KL(q(z|x)||p(z)), \tag{5.5}$$

where KL is the Kullback-Leibler distance between p and q, $q(z|x)$ is the output of the convolutional based encoder portion of our VAE and $p(x|z)$ is the output from the decoder section.

We make use of a Gaussian prior as the distribution for $p(z)$. Gaussian distribution might not be the most suitable prior, but we follow the prior work from [108]. Figure 5.4 shows the the general layout of our VAE model.

Implementation Details We utilise a convolutional encoder for our VAE model and a transpose convolution based decoder. The encoder consists of a 1D convolutional layer, with batch normalisation and leaky ReLU activation function ($slope = 0.2$), followed by max pooling. This common learned feature representation is passed into two separate linear layers which learn the μ and σ used to parametrise the Gaussian and generate z. The decoder architecture comprises three stacked 1D transpose convolutional layers all using leaky ReLU ($slope = 0.2$) to perform the reconstruction from z to create $\hat{x}$.

As with the GAN models, all implementation and training is carried out in *PyTorch* [181], with Adam [107] used as the optimization algorithm (momentum $\beta_1 = 0.1$, $\beta_2 = 0.999$, initial learning rate $\alpha = 0.0001$).

5.4 Experimental Setup

This section will detail the setup of our experimental evaluation, including introducing the empirical datasets used, detailing how we generate new data samples from our generative models and our procedure for evaluating the quality of the generated data.

5.4.1 Empirical SSVEP Dry-EEG Datasets

We make use of two empirical SSVEP dry-EEG datasets where the collection procedure for this dataset is detailed in Figure 5.5. This empirical data is used as a way of validating that the generated data is realistic enough to be used to improve the performance of a classification model. In this dataset, objects are detected in a video sequence using a pre-trained object detection model [138]. The detected objects are then flickered by rendering black/white polygon boxes on top of the objects with display frequency modulations of 10, 12 and 15 Hz to create the frequencies, common to both datasets, which we attempt to detect via the SSVEP paradigm as in Chapter 4. Our two empirical datasets are detailed below:

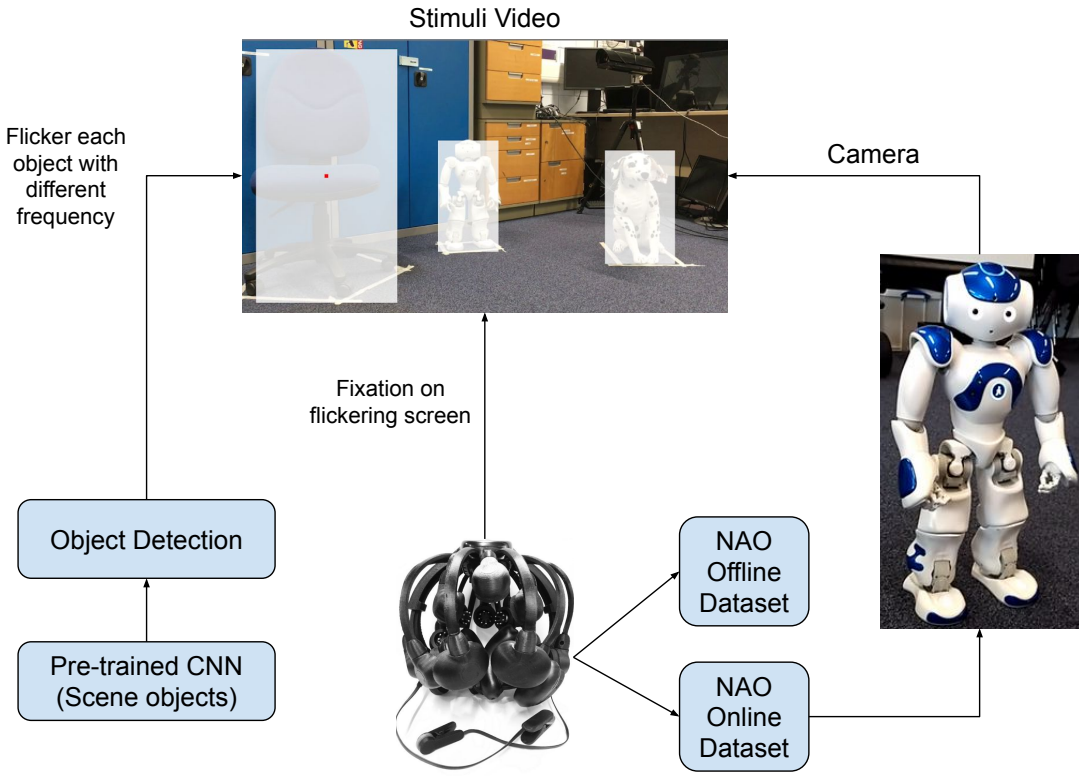


Figure 5.5: Flowchart detailing the recording procedure for the *Online* and *Offline* SSVEP EEG Datasets. The highlighted region of the figure shows the humanoid robot used to create the *Online dataset*.

- *Video-Stimuli Dataset:* This dataset is collected from an offline experiment (shown in Figure 5.1) and is used as a basis for our generative models to learn the distribution of the SSVEP data. This dataset comprises 50 unique samples recorded for each of the three classes, taken from one subject (S01) performing the SSVEP task by looking at objects detected in a pre-recorded video sequence using a web camera. Apart from using different video sequence, the procedure of collecting this dataset is similar as in Chapter 4.4.4.

- *NAO Dataset:* This dataset comprises data from an experiment containing both offline and online elements (highlighted in Figure 5.5) which we collected and fully detailed in Chapter 4. The offline portion of the data contains 50 unique samples for each of the three classes taken from three subjects (S01, S02, S03) performing the SSVEP task by looking

at objects detected in a pre-recorded video sequence from a humanoid robot. The online portion of the data contains 30 samples per class taken from the same three subjects when navigating the robot in real time.

5.4.2 Synthetic Data Generation Procedure

Each of the generative models are trained upon data taken only from the *Video-Stimuli dataset*, detailed in Section 5.4.1. This will allow us to explore if data generated under one condition, can be used to improve the classification of data recorded under another. Once training is complete, the models are used to generate entirely new and synthetic datasets, with one dataset being made for each model. A potentially unlimited amount of data could theoretically be generated from our models. For our experiments, we generate 500 unique samples, each three seconds in length, for each of the three SSVEP frequencies present in the original *Video-Stimuli dataset*.

5.4.3 Classification Procedure

In our experimental evaluation, we investigate exploiting the generated data to improve the classification accuracy on the *NAO dataset*. To perform this classification, we employ a model based on the SSVEP Convolutional Unit (SCU) architecture based on Chapter 3 (Figure 5.6) with some modification. Based on [187], we reduce the number of electrodes use for the SSVEP classification training from 9 sensor channel to 2 occipital cortex channel (O1 and O2) and a frontal cortex (Fz) as data reference channel. Prior work in Chapter 3 has shown this model to outperform traditional approaches and even time-series specific models like Recurrent Neural Networks (RNN) when classifying SSVEP EEG signals. The classification model used for all experiments comprises of 1D convolutional layers, with batch normalization and max pooling. We first pre-process the EEG channels by referencing the data, applying a bandpass filter between 9 to 60 Hz, a notch filter at 50 Hz and finally normalizing the data between 0 and 1.

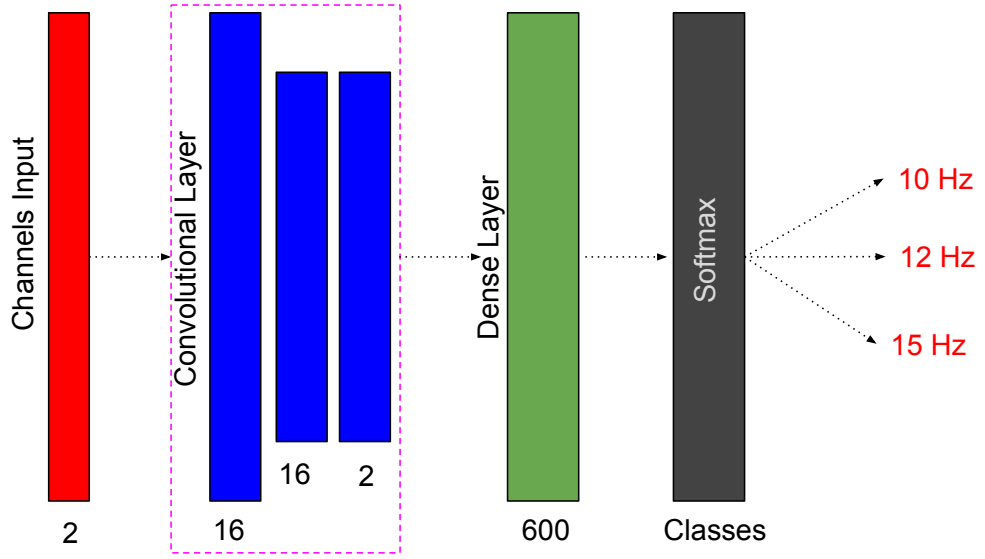


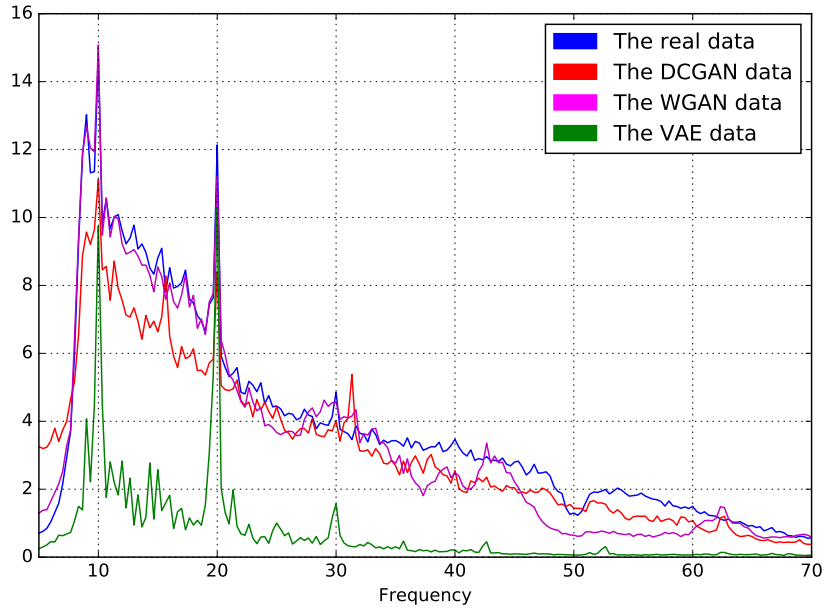
Figure 5.6: The SCU architecture used to classify the EEG signals.

5.5 Results

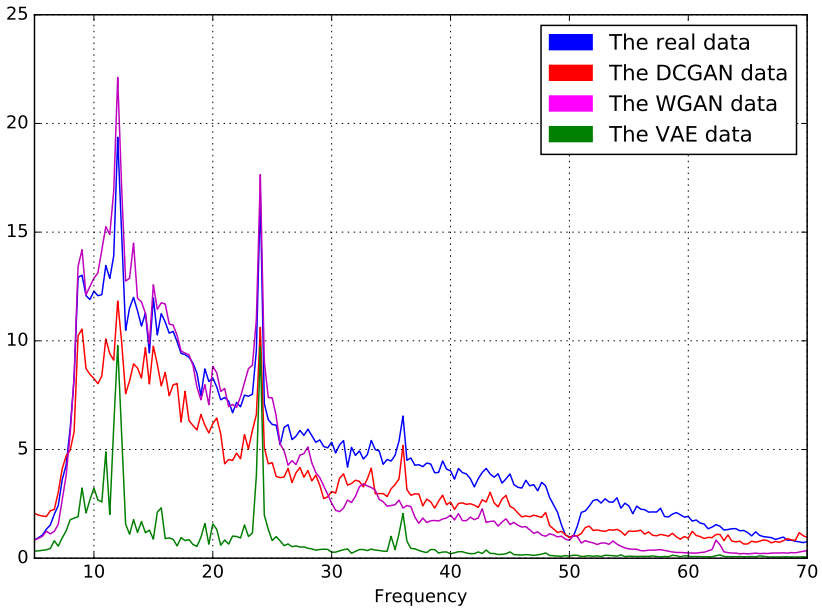
Our experimental evaluation is designed to firstly assess if it is indeed possible to generate realistic EEG signals containing SSVEP signals and secondly, if this synthetic data can be used in a variety of ways to improve the classification accuracy on other real-world empirical datasets. All the classification results presented are the mean results from five different random seeds in order to test robustness and each experiment uses identical train/test splits for all runs to allow for direct comparisons to be made.

5.5.1 Data Visualization via Fast Fourier Transforms

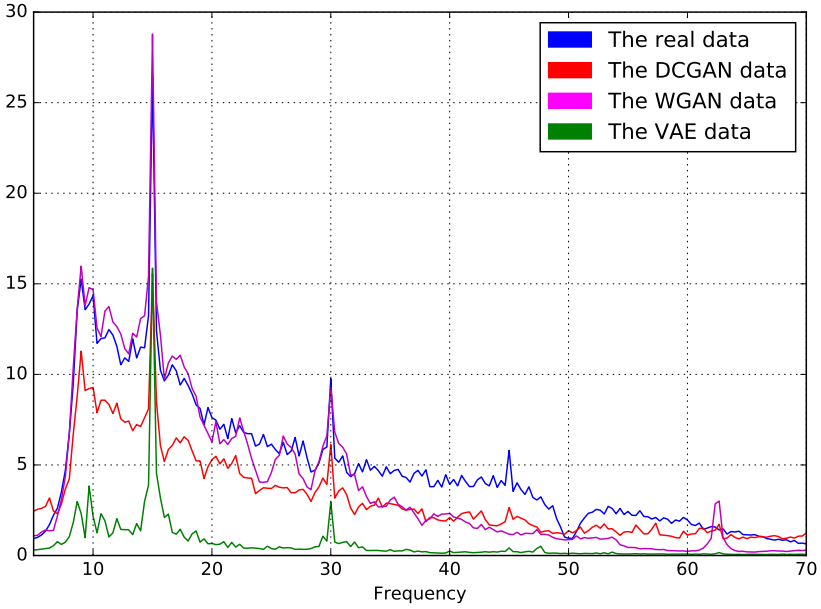
SSVEP data has the phenomenon of frequency tagging, where the primary visual areas in human cortical oscillates to match the frequency of the fluctuating sinusoidal cycle of the SSVEP stimuli presented to the subject [7]. EEG signals recorded during an SSVEP task will contain the target frequency clearly identifiable in the frequency domain [136].



(a) SSVEP Signal at 10Hz.



(b) SSVEP Signal at 12Hz.



(c) SSVEP Signal at 15Hz.

Figure 5.7: Comparing real and synthetic data from the generative models. Synthetic data clearly displays the characteristic SSVEP frequency peaks at the same frequencies as those observed in the real data.

To validate that our generative models are indeed producing viable SSVEP signals, we visualise both the real and synthetic data via the Fast Fourier Transform (FFT) to decompose the EEG signals into the frequency domain. Figure 5.7 displays the frequency plot of the new synthetic data generated from our DC-GAN, WGAN and VAE compared against the empirical data. As can be seen, all models accurately capture the characteristic SSVEP peaks at the target frequency and associated harmonics [194], with the VAE producing signals with a comparatively lower amplitude. This result is very encouraging as it strongly suggests generative models can produce realistic SSVEP dry-EEG signals.

5.5.2 Mixed Real and Synthetic Data Classification

To evaluate the applicability of synthetic EEG signals being used to improve classification results, we combine the synthetic (using 30 samples per class) and real data (30 samples per

class) into a single training set from which the classifier (see Section 5.4.3 for details) is trained. Table 5.1 displays the results, with values given per and across all subjects, where the baseline is no synthetic data included in the training set. The other three method entries all have synthetic data added generated from the corresponding approach. It is also interesting to note that no single generative models data source performs the best for each subject individually, although the best performance across each subject are all within the standard error of each other, but there is no clear winner.

Method	S01	S02	S03	Across Subjects
Baseline	0.91 ± 0.04	0.87 ± 0.10	0.84 ± 0.03	0.69 ± 0.03
DC-GAN	0.86 ± 0.01	0.80 ± 0.04	**0.89** ± 0.03	**0.72 ± 0.03**
WGAN	**0.93 ± 0.04**	0.90 ± 0.04	0.87 ± 0.02	0.71 ± 0.04
VAE	0.92 ± 0.03	**0.90 ± 0.03**	0.79 ± 0.03	0.67 ± 0.02

Table 5.1: Classification test accuracy using generated and real-world data used to train the classifier. The baseline result is the classification of only real data.

Having a single classification model perform well across subjects is known to be highly challenging [56], yet possible [15]. Table 5.1 illustrates how the inclusion of synthetic data within the training set can positively influence generalisation capabilities across subjects, which is an important result and of significant value in real-world applications. It is also interesting to note that although the synthetic data was generated from a different stimuli and sessions, its inclusion is still capable of improving the classification accuracy.

5.5.3 Classification with Pre-Training

To further explore the usefulness of the synthetic data, we pre-train our classifier using only synthetic data (with 500 samples generated per class) and then further fine-tune the model using the real data as in Figure 5.8. The performance on the testing set is presented in Table 5.2, where the baseline is no pre-training using synthetic data. The other three method entries all have been pre-trained using synthetic data generated from the corresponding approach. Figure 5.9 highlights how the loss values for this task vary over training epochs. It can be seen that models pre-trained using the synthetic data converge faster and to a lower overall loss value.

Both Table 5.2 and Figure 5.9 demonstrate the ability to use the synthetic data to pre-train the network and improve the test accuracy on real data - another key result. This demonstrates

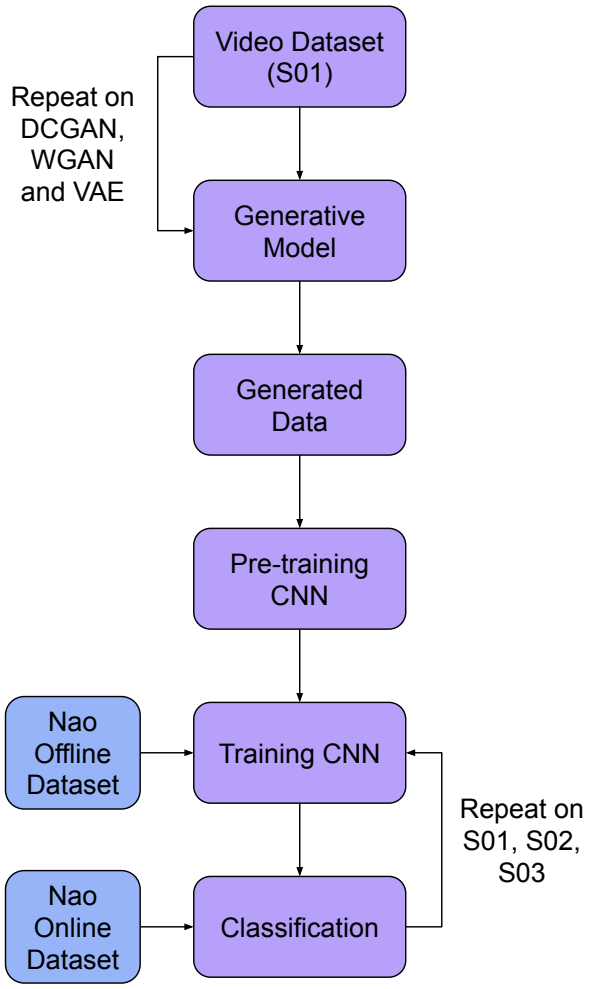


Figure 5.8: Flowchart for Pre-training classification experiment.

the possibility of achieving higher classification accuracy using a smaller training set of real-world data, which resolves one of the most important challenges associated with SSVEP signal classification, unavailability of large datasets.

Additional experiments were also conducted using varying quantities of synthetic data during pre-training. As seen in Figure 5.10, the models perform better on average when 500 synthetic data samples are used in the pre-training stage. This is primarily due to the subject and session

variant nature of dry-EEG signals, since the models can easily over-fit to the distribution of the synthetic data in the presence of excessive pre-training. Furthermore, Figure 5.11 demonstrates how the training time can rapidly increase when the size of the synthetic pre-training dataset grows too large [4]. Accordingly, the use of 500 synthetic data points empirically offers an optimal trade-off between improved performance and tractable training time in this instance.

Method	S01	S02	S03	Across Subjects
Baseline	0.91 ± 0.04	0.87 ± 0.10	0.84 ± 0.03	0.69 ± 0.03
DC-GAN	**0.97 ± 0.03**	**0.93 ± 0.03**	**0.87 ± 0.01**	0.70 ± 0.02
WGAN	0.93 ± 0.03	0.90 ± 0.04	**0.87 ± 0.03**	0.72 ± 0.03
VAE	0.93 ± 0.03	0.92 ± 0.06	**0.87 ± 0.05**	**0.73 ± 0.03**

Table 5.2: Accuracy for test classification using synthetic data for pre-training stage. The baseline contains no pre-training.

5.5.4 Cross-Subject Generalisation

As mentioned in Chapter 5.1, SSVEP classification models often exhibit poor generalisation performance across different subjects or experiential data collecting sessions [140, 226]. In previous workin Chapter 3.5.4, we successfully classified an unseen subject with no additional training required on that subject.

In this section, we also attempt to demonstrate how pre-training on synthetic data can enhance the cross-subject generalisation capabilities of the model, i.e., the ability to classify data captured from one subject using a model trained on data captured from another. We train the generative models on S0X NAO Offline data (from the *Nao Dataset*), pre-train the classifier using 500 samples of synthetic data output by said generative models, train the classifier on data from a different subject (S0Y) and finally test on online data from S0X from the *Nao Dataset* (Figure 5.12). As commonly seen within the literature [142, 56], one of the most important challenges in EEG-based research is the properties of EEG signals that vary from one subject to another as signal features can be specific to individual subjects.

Table 5.3 demonstrates how the baseline model performs poorly when trained using data collected from S0Y but tested on data from S0X, pointing to the lack of model Generalisability

[4] The training code has been written using Pytorch autograd package and performed on a Nvidia GeForce GTX 1060 GPU

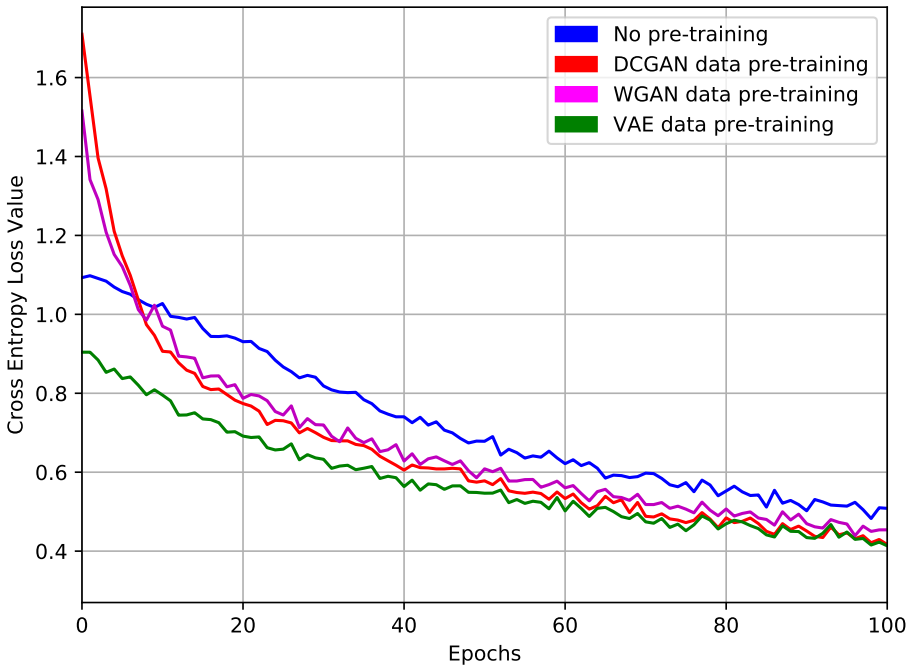
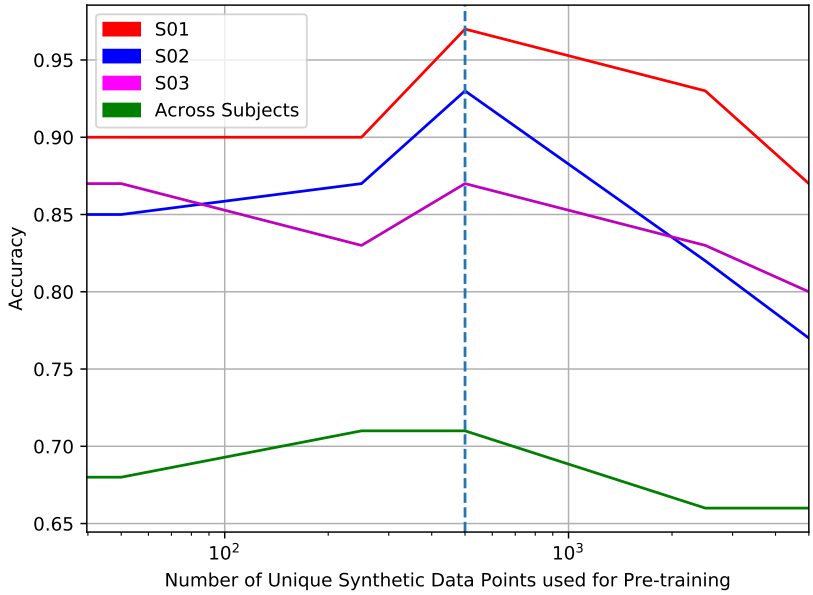


Figure 5.9: Convergence of the Cross-Entropy value plotted over training epochs for models with and without pre-training on synthetic data.

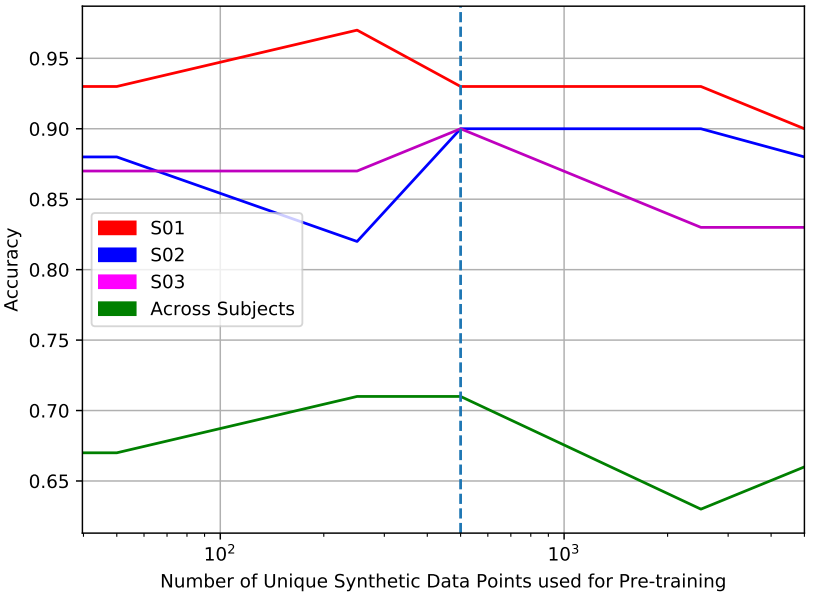
when EEG signals are used. However, by pre-training the classifier on the generated data using NAO offline data from S0X, we see a large improvement in classification accuracy - especially when data generated by the VAE is used. This is a significant observation in the BCI research domain leading to a conjecture that, with further improvement in the results, we can eliminate the requirement for per-subject, per-session calibration for online applications.

5.6 Conclusion

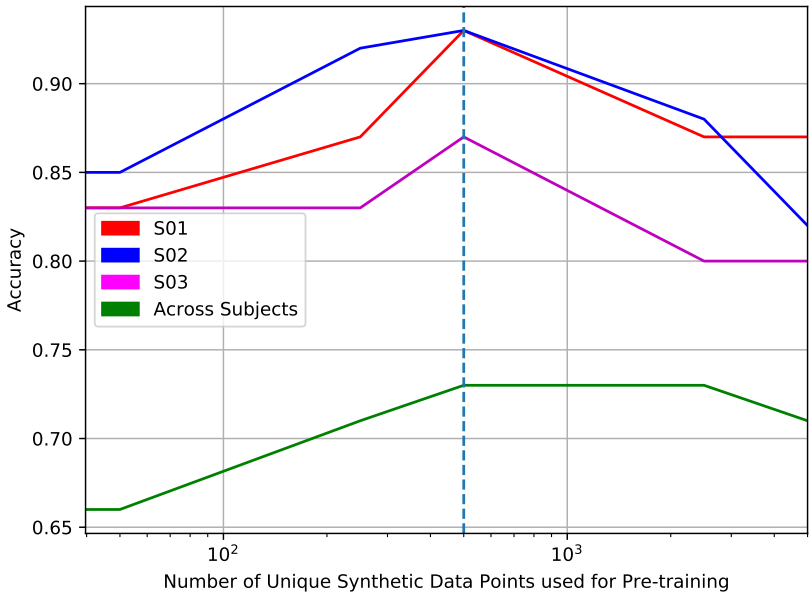
In this chapter, we exploit recent advances made in neural-based generative models to explore the potential benefits they can offer within the context of SSVEP classification models trained on EEG signals. Since data acquisition within real-world scenarios suffers from a variety of challenges, using synthetically generated EEG signals can prove highly beneficial in improving



(a) DC-GAN



(b) WGAN



(c) VAE

Figure 5.10: Test accuracy when varying the volume of synthetic data used for pre-training. Doted line indicates 500 samples.

Method	S01	S02	S03	Mean
Baseline	0.35	0.54	0.42	0.45 ± 0.08
DC-GAN	0.42	0.71	0.57	0.57 ± 0.12
WGAN	0.40	0.60	0.50	0.56 ± 0.13
VAE	0.70	0.90	0.85	**0.82 ± 0.08**

Table 5.3: Classification test accuracy for cross-subject Generalisation (see Figure 5.12) via pre-training using synthetic data. The baseline does not include a pre-training stage.

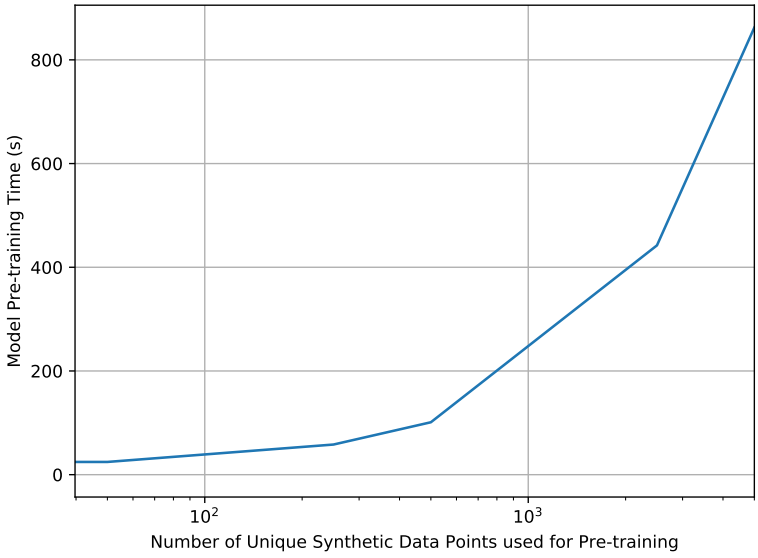


Figure 5.11: Pre-training time

the accuracy, convergence rate and generalisation capabilities of any model trained to classify EEG signals. In comparison to the prior work [253, 191], we generate EEG signals directly in signal space via end-to-end training instead of first transforming the signals into different domains for SSVEP classification. We generate synthetic EEG signals using three state-of-the-art generative models - a Generative Adversarial Network, a Wasserstein Generative Adversarial Network and a Variational Auto-Encoder. Extensive evaluations demonstrate the efficacy of the synthetic data generated by said models across multiple experimental setups, with the inclusion of the generated data always improving the results.

Although this work provides solutions in generating realistic synthetic EEG signals, the results presented in this chapter are based on a dataset comprised of signals taken from just three subjects. The synthetic signals generated from various generative models are realistic however, it is still impossible for the SSVEP classifier to generalise the use of only synthetic signals across real EEG signals. Here, in addition to mixing the real and synthetic data where the number of data used is equal, we take the pre-training approach where we pre-train on the larger amount of synthetic data before further fine-tuning the network on the real-data. As we highlight in Section 1.1 and Section 2.4, SSVEP classification models often exhibit poor generalisation performance across different subjects or experiential data collecting session. In this chapter, we demonstrate

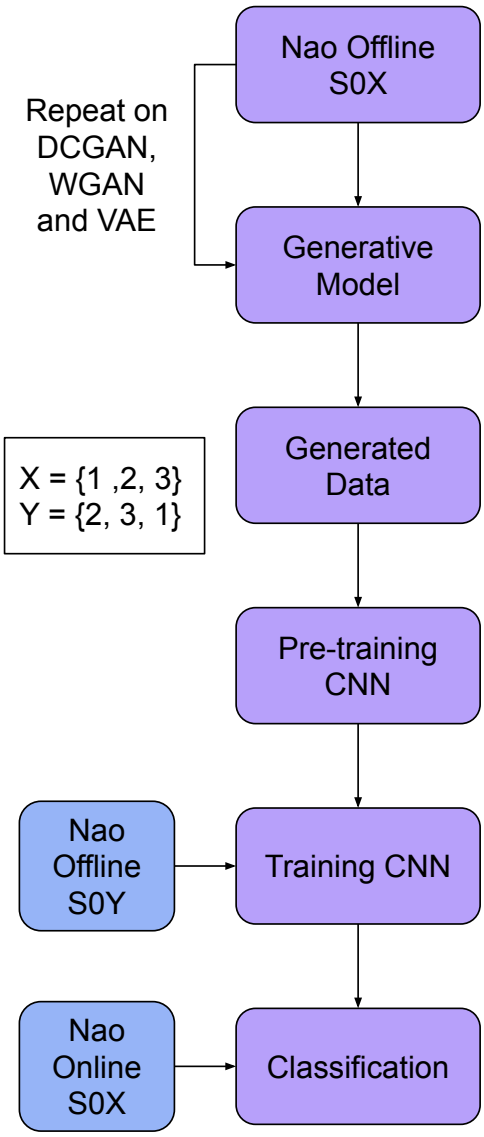


Figure 5.12: Flowchart for Training on data from different subjects.

that without synthetic signals trained on Subject X being included, the SSVEP classification network struggles to classify Subject X when the network has just been trained on Subject Y. Although we manage to boost the performance of the SSVEP classification, we still have not

managed to completely solve the unseen subject problem.

Future work will investigate the influence of having a single model to generate signals from different frequency classes. This could address the limited data problem by essentially exposing the generative model to three times the amount of data, thus providing a bigger dataset (Section 4.4.4). Furthermore, we also plan to study the ability of the generative model to eliminate subject information whilst generating a number of realistic synthetic EEG signals. Finally, a larger dataset should be used in order to create further justification that it is possible for a GAN to generate realistic synthetic EEG signals that can be used in downstream BCI applications.

Chapter 6

Synthetic EEG Signals via Subject Invariant GAN

Moving forward with remarkable strides have been made by the capabilities of modern neural-based generative models, in this chapter, we propose a novel means of generating highly-realistic synthetic EEG data invariant to any subject, session or other environmental conditions. Focusing on a Steady State Visual Evoked Potential (SSVEP) classification system, we propose a novel means of generating highly-realistic synthetic EEG data invariant to any subject, session or other environmental conditions. Our approach, entitled the Subject Invariant SSVEP Generative Adversarial Network (SIS-GAN), produces synthetic EEG data from multiple SSVEP classes using a single network. Additionally, by taking advantage of a fixed-weight pre-trained subject classification network, we ensure that our generative model remains agnostic to subject-specific features and thus produces subject-invariant data that can be applied to new previously unseen subjects. Our extensive experimental evaluation demonstrates the efficacy of our synthetic data, leading to superior performance, with improvements of up to 16% in zero-calibration classification tasks when trained using our subject-invariant synthetic EEG signals.

The work presented in this chapter has been published in the following peer-reviewed publication:

Nik Khadijah Nik Aznan, Amir Atapour-Abarghouei, Stephen Bonner, Jason D Connolly, and Toby P Breckon. Leveraging Synthetic Subject Invariant EEG Signals for Zero Calibration BCI. In International Conference on Pattern Recognition. *IEEE, 2020*

6.1 Introduction

In this chapter, we predominantly focus on the SSVEP paradigm as the visual stimuli, which evokes a neurophysiological response in a human viewing a frequency-based visual stimulus [194], making it a prime candidate for use in teleoperation tasks. However, the resulting EEG signals are highly subject and session dependent, so much so that the signals contain patterns unique to specific subjects and have even been used as a biometric [67, 11, 248]. This leads to severe calibration requirements for any EEG-based BCI application. Calibration is needed at the beginning of any task to account for specific subject and session conditions [112, 142] as illustrated in Figure 6.1.

Many EEG-based applications, for example teleoperation, could be used as assistive technologies if not for their subject and session dependency [142], that restricts real-world practical applications. This extreme subject dependence is demonstrated in Figure 6.2, where a confusion matrix is seen for subject-biometric classification. A simple convolutional neural network easily manages to classify SSVEP-based EEG signals from nine subjects, indicating the existence of a distinctive subject bias, where each subject signal contains easily-distinguishable biometric features. In this paper, we investigate the possibility of leveraging recent advances in deep machine learning to generate synthetic EEG signals which are subject invariant and hence can be used to enable the training of generic EEG signal decoders for previously unseen subjects. Recently, neural-based generative models, such as Generative Adversarial Networks (GAN) [82], have been shown to be capable of capturing the key elements of a given dataset by learning a hidden structure from its underlying distribution to generate new data samples within the same distribution. Previous work in Chapter 5 has even demonstrated significant improvements in performance if synthetic EEG signals generated by a GAN are used for pre-training an SSVEP classification network.

However, as EEG signals are highly subject-specific, the issue of subject bias has to be taken into consideration when dealing with such data in a learning-based approach [104, 244]. Machine

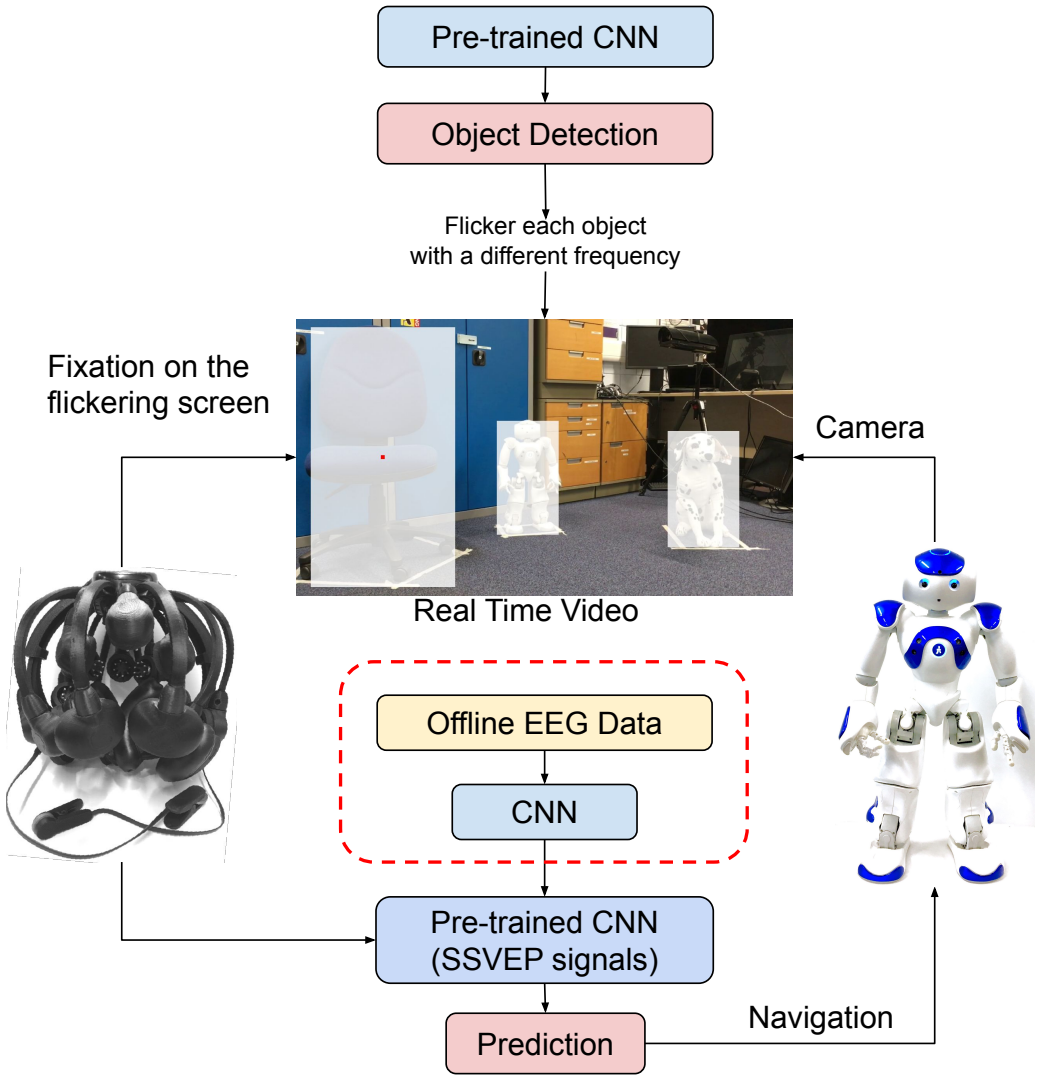


Figure 6.1: Teleoperation task with the required calibration stage identified in the red dotted block.

learning, in general, is fraught with data-specific issues as models often have a tendency to overfit to unapparent or unobservable nuances of a given dataset in order to maximise task performance. Particular features in a dataset can often drive the resultant model distribution toward a certain

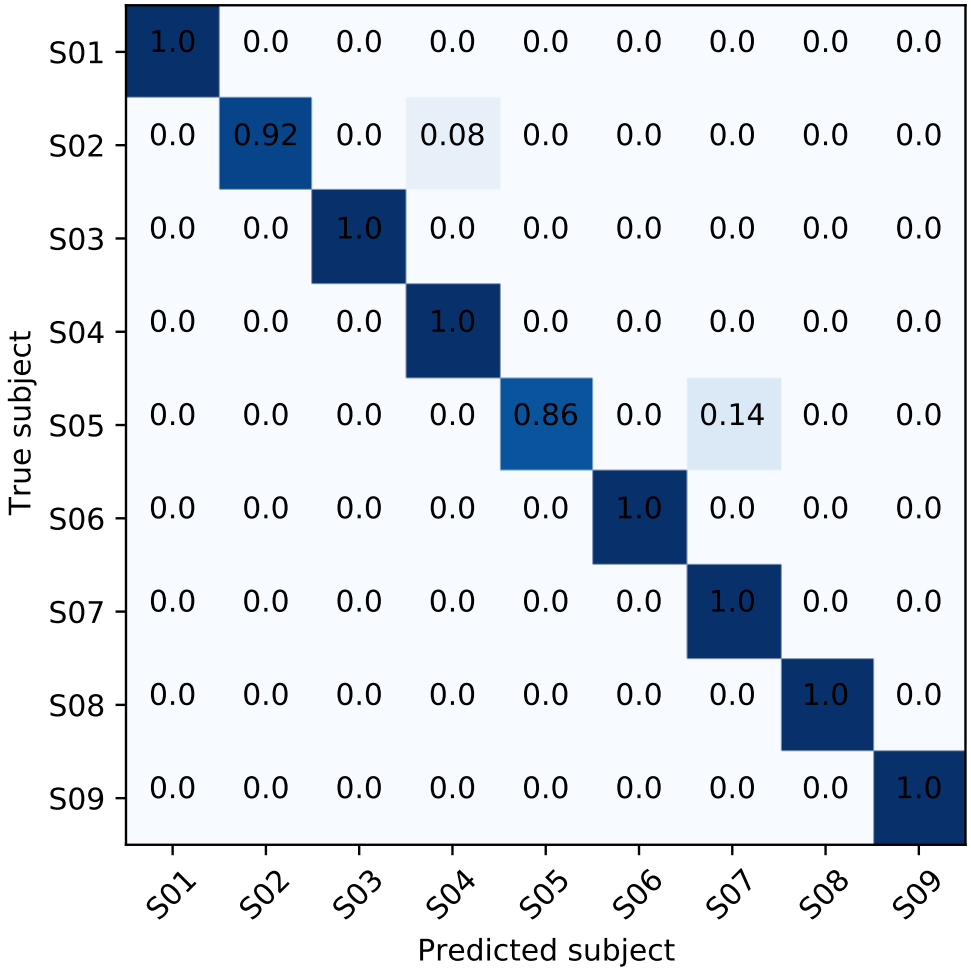


Figure 6.2: Confusion matrix for subject classification accuracy (Maximal result being accuracy = 1.0 on the diagonal).

direction resulting in bias and inferior performance on previously unseen data.

In this chapter, we exploit the capabilities of a Subject Invariant SSVEP Generative Adversarial Network (SIS-GAN) to produce not only realistic synthetic EEG signals but also to eliminate subject-specific features in order to boost performance on subsequent tasks. To demonstrate this, we choose to show results on an SSVEP-based teleoperation task, although

potential applications are not limited just to this. In this cross-task setting, the data utilised to train our generative model are from different subjects and different capture sessions that remain unseen for the downstream SSVEP classifier. The generated signals are used to classify unseen subjects from the online EEG signals from Chapter 4 where three subjects navigate a humanoid robot using variable position and size SSVEP stimuli. Our experimental results (Section 6.5) demonstrate the efficacy of our SIS-GAN approach.

In summary, the major contributions of this paper are:

- *Zero Calibration* - Our novel model architecture results in synthetically-generated subject-invariant EEG signals suitable for real-time BCI applications with near zero calibration.
- *Cross-Task Results* - Downstream models trained on data generated by our SIS-GAN approach are shown to perform well when classifying Online signals having only seen Offline signals during training.
- *Realistic Synthetic Data* - We use our deep learning based generative model to create highly *realistic synthetic EEG signals* shown to offer superior performance for downstream signal decoding tasks.
- *SSVEP Classification for Unseen Subjects* - Our use of enforced subject invariance, via our SIS-GAN architecture, results in models capable of accurate SSVEP classification for unseen subjects with improvements of up to 16% against contemporary state-of-the-art approaches.

To validate our claims and to aid reproducibility, we release our implementation of the proposed GAN models here - https://github.com/nikk-nikaznan/Subject-Invariant-SSVEP-GAN

6.2 Related Work

Calibration issues have long plagued BCI applications, as calibration is required at the beginning of every session [112], even for the same subjects doing the same tasks, leading to subject fatigue, data variability and hence adverse effect upon signal quality. This is not only time-consuming and tedious but can also cause subject fatigue which then affects the quality

of the data. The authors in [112] introduce the zero training method where common spatial patterns are used to learn features by transferring knowledge from previous sessions. While the approach can lead to a reduction in task time without jeopardising accuracy, it is subject-specific as the objective is not to eliminate calibration for new unseen subject.

In [125], the authors propose a multi-source transfer learning method to classify a new subject by learning Style Transfer Mapping (STM) between new and existing subjects. They use the prior knowledge from existing subjects by calculating the minimum difference between the existing and new subjects. They require a few calibration sessions for each new subject and need to learn the ensemble weight by classifying data from the calibration sessions using each individual classifier. This ensemble weight is used to integrate all the individual classifiers transformed by STM to function in the presence of new unseen subjects.

Our previous work as in Chapter 3 demonstrates the capabilities of a classification model generalising to entirely new unseen subjects using a convolutional neural network (CNN) trained on three subjects to perform a simple SSVEP-based task without any additional training needed for a new subject. However, reductions in the classification accuracy could be noticed due to the lack of features unique to the unseen new subject.

In this chapter, we propose a novel data generation framework taking advantage of generative models to produce subject-invariant synthetic EEG data that could significantly enhance any downstream training procedure. The use of generative models such as Generative Adversarial Networks (GAN) within the realm of BCI, however, is not entirely unknown. For instance, in Chapter 5, we previously investigated the possibility of generating new synthetic EEG data containing SSVEP information using a selection of customised unsupervised neural-based generative models trained on a limited quantity of EEG data from different subjects.

Similarly, in [89], the authors utilise a Wasserstein GAN [9] trained on single-channel EEG-based motor imagery data with four evaluation metrics: inception score, Frechet inception, Euclidean and sliced Wasserstein distance. The work in [53] also uses a Wasserstein GAN but to generate high-resolution EEG-based motor imagery data by interpolating one channel to another using low-resolution data.

By using synthetic data generated from Recurrent GAN (RGAN), in [1] the authors augment the dataset of motor imagery dataset with synthetic data of different augmentation sizes. Even

though the classification accuracy using RGAN augmentation data outperform Variational Auto-Encoder and Auto-Encoder, this work still not improving the classification accuracy of using just real data.

The work in [98] proposes an EEG-based zero-shot learning GAN to classify ten unseen EEG labels. A GRU-based encoder is first used to extract the real EEG features, and a conditioned Wasserstein GAN is subsequently employed to generate fake EEG features using *word2vec* [158] as a semantic embedding input. During the training process, word embedding features of unseen class labels are used to generate fake EEG features.

Within the wider scope of machine learning, there has recently been advances towards removing undesirable bias towards specific data features, which is of significant relevance to our work. For instance, in [104], a new training procedure is proposed wherein the network is trained with biased data but is incentivised not to learn some specified target bias and is therefore able to perform well on unbiased test data. They formulate a regularisation loss to minimise the mutual information between the features and the bias.

The work of [244] proposes a Fairness-aware Generative Adversarial Network (FairGAN) where the network generates data without having any information about certain 'protected attributes'. The approach has a generator conditioned on the protected attributes and two discriminators to identify whether the data is real or fake and to ensure that the generated fake samples do not contain any of the undesirable features.

These recent works [104, 244] on identifying the key features show potential for achieving subject invariant EEG signals. Perhaps by learning the bias features related to subject and training the generator to produce signals without the said features, we may be able to leverage it in the subject variant misclassification problem.

Inspired by these advances in unbiased training frameworks, in this chapter, we investigate the possibility of generating synthetic EEG signals using SIS-GAN subsequently used to classify unseen subjects in a downstream EEG-based teleoperation task. While earlier work [112, 125, 98] discuss the difficulties often faced with the calibration stage, no other work, to date, has managed to classify completely unseen subjects without re-training, fine-tuning and having a large dataset. In this vein, we also propose a novel adversarial generative model (Section 6.3.3) capable of producing highly valuable subject-invariant EEG data that could be used to augment the training data needed for EEG-based models, leading to a significant boost in performance.

6.3 Proposed Approach

In this section, we outline our proposed Subject Invariant SSVEP Generative Adversarial Network (SIS-GAN) approach to generating subject invariant EEG signals containing SSVEP information. The SIS-GAN model architecture is presented in Figure 6.5 and is comprised of four primary components: a Generator network and its corresponding Discriminator (Section 6.3.1), an Auxiliary classification network (Section 6.3.2) and a pre-trained Subject-biometric classifier (Section 6.3.3), enabling the generation of subject invariant signal samples.

6.3.1 Generator and Discriminator Networks

The backbone of our approach for EEG signal generation is inspired by prior work on producing synthetic images [192]. In this Generative Adversarial Network (GAN) setup, the generator, G, receives as its input random noise vectors, z, sampled from a Gaussian distribution (as in Figure 6.3).

This generator produces fake data samples ($\tilde{x} = G(z)$) at every iteration, which along with real data samples, x, are used as the input to a discriminator D. As the discriminator is trained to classify the data samples as either fake or real, the resulting gradients are successively used to train the generator, leading to higher quality fake samples to the point where they become indistinguishable from the real ones. The training objective consequently relies on the competition between the generator and the discriminator following the minimax objective [82]:

$$L_D = \min_G \max_D \mathop{\mathbb{E}}_{x \sim \mathbb{P}_r} [log(D(x))] + \mathop{\mathbb{E}}_{\tilde{x} \sim \mathbb{P}_g} [log(1 - D(\tilde{x}))], \tag{6.1}$$

where $\mathbb{P}_r$ is the real data distribution, $\mathbb{P}_g$ the model distribution defined by $\tilde{x} = G(z), z \sim p(z)$, and z the random noise vector used as the input to the generator.

It is important to note that a vanilla GAN would only be capable of capturing the underlying distribution of and thus generating one particular class of data at a time. Due to this limitation, in the following section, we detail the introduction of a Auxiliary classification component to allow the model to produce data from multiple classes simultaneously.

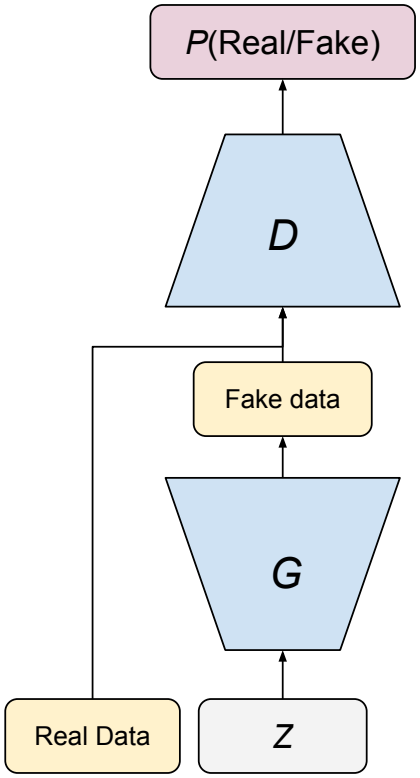


Figure 6.3: The Deep Convolutional Generative Adversarial Network (DC-GAN) architecture, where the Generator (G) is trained to fool the Discriminator (D).

6.3.2 Auxiliary Classifier Network

An interesting alternative to training separate models that would generate data on a class by class basis is to have one model capable of producing data from all classes as required. In an auxiliary Classifier GAN (AC-GAN), [176], the input to the generator is not only a random noise vector but also a class label for the generated output (Figure 6.4).

The generator is consequently trained to produce fake data samples from a model distribution similar to the real data distribution for each specific class label. The discriminator will essentially identify whether the generated data is real or fake while at the same time classifying which class

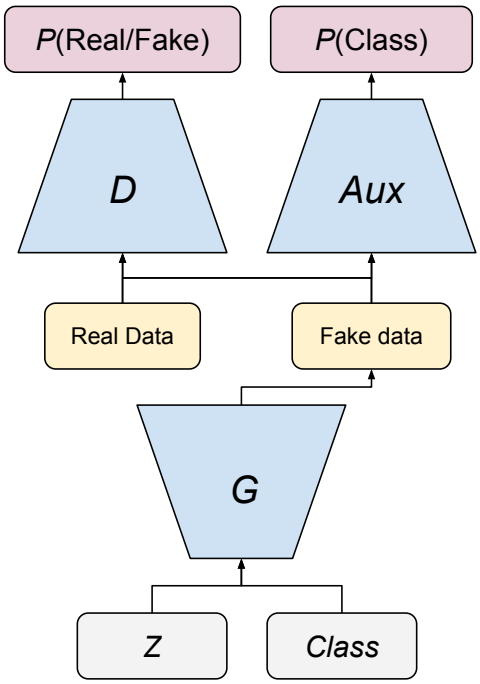


Figure 6.4: The Auxiliary Classifier Generative Adversarial Network (AC-GAN) architecture, where the Generator (G) is trained to produce data which can fool the Discriminator (D) and to produce data which can be classified as belonging to a certain class by the Auxiliary network (Aux).

the generated data sample belongs to. Not unlike [192], the discriminator and the generator networks in an AC-GAN approach are trained to maximise each corresponding objective function but with two associated loss components: loss of the source (real or fake) and loss of the class label. As such, the use of an AC-GAN enables us to train a single model capable of generating EEG signals for all subjects and all three SSVEP classes.

As the auxiliary component is fundamentally a classification task, Cross-entropy is used as the loss function. Cross-entropy functions by measuring the probability between the predicted class (x) value to the actual value (y):

$$L_A = -\Sigma(x \log y). \tag{6.2}$$

6.3.3 Creating Subject Invariant EEG Signals

Inspired by recent work on bias removal [244, 104], we propose a novel synthetic EEG signal generation pipeline capable of producing subject-invariant signals that are unbiased representative samples for any unseen subject. Knowing the target bias, [104] proposes a network targeting specific features to penalise the model when said features become prominent in the model distribution. For instance, a network can easily be biased towards using colour information as a cue while colour has indeed no relevance to the actual task, such as handwritten digit classification. Therefore, by specifically targeting RGB information, the bias can be eliminated, leading to a more robust representation learning. However, in our dataset, there is no specific feature that can be identified as the culprit when it comes to introducing subject bias. EEG data is known to be complex, with every subject and session being uniquely identifiable - a phenomenon still not fully understood [112, 125].

Consequently, we design our training process in such a manner that the network is rewarded when it learns features that are common across all subjects and is penalised when subject-specific features can lead to subject-biometric classification. To achieve this, we introduce a pre-trained frozen subject-biometric classification network (shown in the upper-right of Figure 6.5), charged with classifying which subject the generated EEG signal biometrically belongs to. It is important to reiterate that this classification network is frozen and the gradients from its loss function are only used to train the generator and not the network itself. Since a correct classification would be undesirable, the gradients from this pre-trained frozen network can be used to penalise the generator, thus pushing it towards generating more generic subject-invariant outputs. More formally, the subject invariant loss component can be viewed as the following objective:

$$L_S = \underset{\hat{y}}{\operatorname{argmax}}\, S(\hat{y}|x), \tag{6.3}$$

where S is the pre-trained frozen subject-biometric classification network, x denotes the generated data and $\hat{y}$ is the predicted probability values for all subjects. Essentially, this component minimises the maximum predicted subject probability which, over multiple training steps, will produce synthetic EEG signals which cannot be correctly classified by the subject network – thus making the data subject invariant.

Figure 6.5 illustrates the overall SIS-GAN architecture and shows how the various components of our approach are connected. The model is able to generate samples for all classes using a single

generator, but is similarly capable of producing subject-invariant EEG signals via the penalty introduced from the subject-biometric classification network. The overall training objective of our model is as follows:

$$L = L_D + \lambda_a L_A + \lambda_s L_S, \tag{6.4}$$

where λ_a and λ_s are used to weight the importance of the auxiliary classification and subject identification components in the overall loss score.

6.3.4 Implementation Details

For the sake of consistency, all the discriminator and classification networks in Sections 6.3.1, 6.3.2 and 6.3.3 follow a similar architecture with four layers containing modules of 1D Convolution, BatchNorm, PReLU and DropOut ($p = 0.5$) followed by a linear layer projecting the resulting features to the number of desired classes. The discriminator in AC-GAN (Section 6.3.2) includes two heads, one for discriminating between real and fake samples and one for auxiliary classification. No max-pooling is used as our experiments show strided convolutions yield better performance.

The architecture of the generator in all models contains five layers of fractionally-strided convolutions with BatchNorm and PReLU. Our experiments with residual connections [91] led to no significant improvements in the results. All implementation is done in PyTorch [181], with Adam [107] providing the optimisation ($\beta_1 = 0.5$, $\beta_2 = 0.999$, $\alpha = 0.0001$). In SIS-GAN (Section 6.3.3), the subject loss in Equation 6.3 is empirically weighted down (by a factor of 0.3 for λ_s) for a more stable training process and improved results.

6.4 Experimental Setup

6.4.1 Datasets

For our experiments, we make use of two sets of Dry-EEG data from none-overlapping subjects, collected under two different conditions: Offline and Online. Both datasets are recorded using Quick-20 dry EEG headset from Cognionics Inc. with 20 dry-EEG sensors. We collect data over the parietal and occipital cortex (P7, P3, Pz, P4, P8, O1 and O2) [133], frontal center (Fz) and A2 reference at 500 Hz sample rate with the stimuli displayed on a 60Hz LCD monitor.

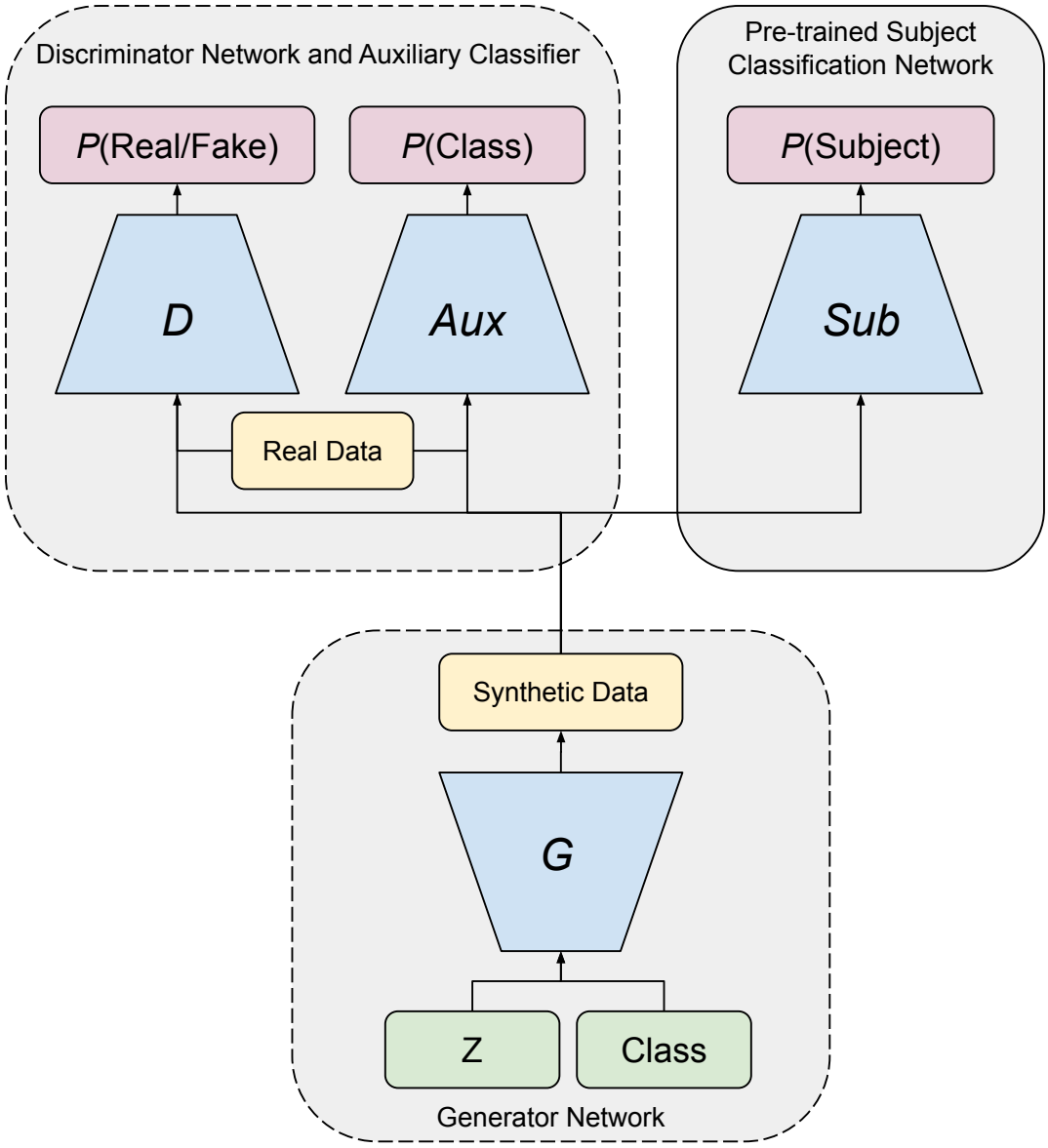


Figure 6.5: Our Subject Invariant SSVEP Generative Adversarial Network. The Generator (G) produces data with subject-specific information removed (Sub) that can fool the Discriminator (D) and is classified as a certain frequency (Aux).

Offline Real Dry-EEG Dataset Data is collected from nine inexperienced subjects (Naïve BCI users) aged 25 to 40 years old from an offline SSVEP experiment (S01 to S09). Cortical signals from the subjects are streamed and recorded using the dry-EEG headset whilst their gaze is fixated on variable SSVEP stimuli as in Chapter 4. The stimuli is created using detected objects in a video sequence. Black and white boxes are rendered over object blocks, thus simulating a flickering effect with display frequency modulations of 10, 12 and 15 Hz to create SSVEP frequencies. Alternately, the interface displays navigational arrows with the same frequency modulations. The objects in the video sequence are captured using the camera onboard our humanoid robot, NAO [84]. We collect 60 samples for each of the three classes per subject.

Online Teleoperation Task Dry-EEG Data The data is from Chapter 4, where they are collected using our experimental setup for real-time robotic teleoperation based on three subjects (T01, T02, T03) . No subject from the Offline dataset is used and all subjects are experienced participants. Subjects are seated in front of a computer screen to navigate a humanoid robot by fixating on real-time on-screen stimuli. The robot faces a scene containing objects which are detected and flickered with different unique frequencies. The decoded SSVEP signals are then used to navigate the humanoid robot towards the subject-selected objects within the calculated robot motion trajectory. The stimuli interface displayed to the subject alternates between flickering objects and the navigational arrows to allow the robot to be navigated when there are no new objects detected within the scene. To enable the use of this data, the raw EEG signals and the ground truth information are all saved from the real-time experiment. The data contains 30 unique samples per subject. We use this dataset in the cross-task validation experiment (Section 6.5.4), where the objective is to learn a model from the Offline data such that it can correctly classify the data in the Online dataset. From the real-time task presented in Chapter 4.5.2, all subjects demonstrate strong statistical real-time performance based on offline training, with the mean accuracy of 90, 87 and 80% respectively.

6.4.2 Evaluation Methodology

The main purpose of this work is to investigate the capability of a generative model to produce realistic synthetic signals that can potentially be used in real-time SSVEP classification to eliminate the calibration stage of a BCI application.

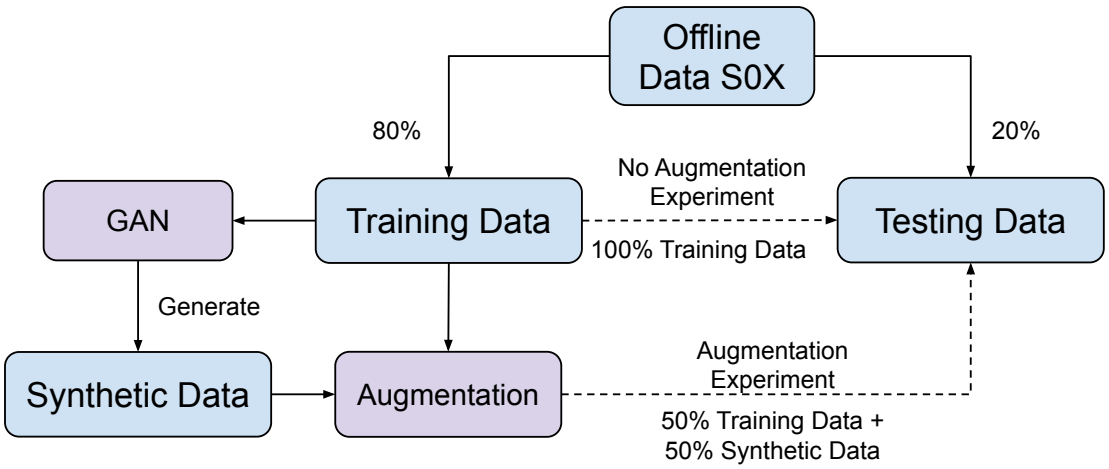


Figure 6.6: Evaluation method for SSVEP classification for single subject, where the training and testing data are from the same subject. Dotted lines indicate SSVEP classification.

The first set of experiments (Section 6.5.1) focus on SSVEP classification for a single subject. Using the Offline dataset, we test the SSVEP classification performance for each subject when the network is trained and tested on the same subject. Results are compared against models trained using realistic synthetic data generated by a Deep Convolutional GAN (DC-GAN) and an AC-GAN. The generative model is evaluated via the augmented dataset containing both real and synthetic data used to classify the testing dataset. Figure 6.6 outlines the experimental procedure. We utilise 20% of the data from Section 6.4.1 for testing. The training dataset is used for training the generative and the SSVEP classification models.

There are two main experiments within this setup: firstly a baseline to see how accurately each subject performs the task and secondly to measure any improvements in performance by augmenting the training dataset with synthetic EEG signals. This is important as most of the time, naïve BCI users do not perform as well as experienced users in a variety of BCI tasks and hence some inter-subject variation is to be expected [196].

It is commonly known for EEG signals to have unique patterns containing specific subject information that leads to difficulty in unseen subject classification [248]. The rest of our experiments focus on the ability of the generated synthetic data to improve the generalisation of the classification model given that there is no prior training on a particular subject. To rigorously evaluate our approach, experiments are carried out using three training datasets:

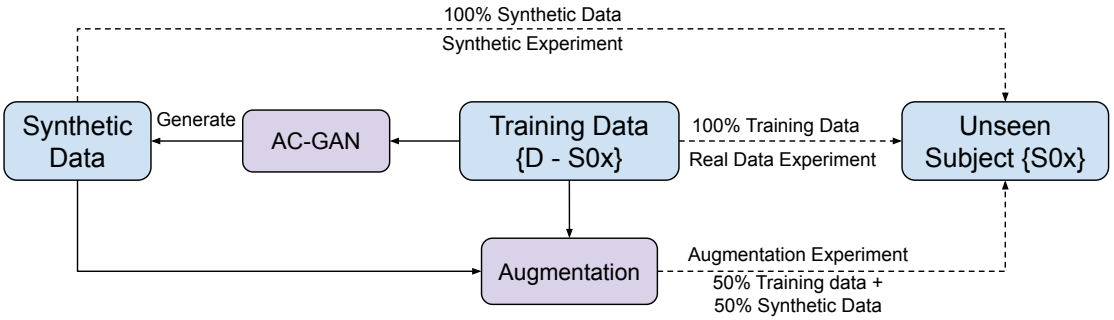
- *Real Training Data* - where the resultant model is trained solely on the real data from the Offline dataset with no synthetic component.

- *Augmented Training Data* - where model training is performed using real data from the Offline dataset mixed with synthetic data from the generative models at a ratio of 50:50.

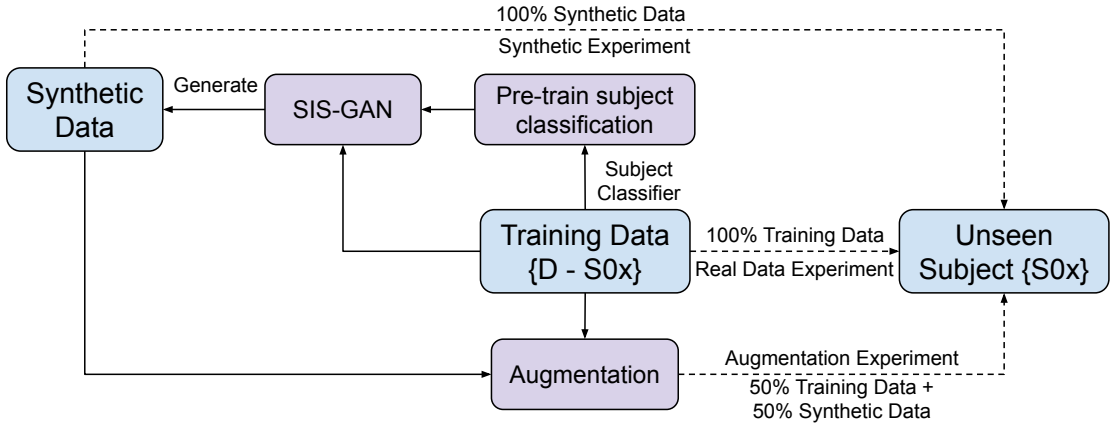
- *Synthetic Training Data* - where model training is performed using only synthetic data from the generative models, with the same number of total samples as the *Augmented Data*.

The second set of experiments (SSVEP Classification for Unseen Subject : Leave-One-Out – Section 6.5.3) likewise focuses on the same Offline dataset {D}. However, instead of training and testing on data captured from one subject, we perform tests on data from one unseen subject {S0x} after training the model on data from all other subjects {D - S0x}. Figure 6.7 shows the experimental procedure, where we are taking the leave-one-out validation approach (i.e. the leave-one-out is {S0x} (test data) and data from all remaining subjects are used as the training data {D - S0x}).

Typically in prior works [142], real-time SSVEP classification results are obtained by calibrating the SSVEP classifier model to the subject during the preliminary task. In our last set of experiments (SSVEP Classification for Unseen Subject : Cross-Task – Section 6.5.4), which are the primary focus of this work, we explore whether this calibration stage can be removed, thus requiring no model retraining before a subject participates in the session. To achieve this, we evaluate the performance of a model on unseen subjects (online task) when the model is pre-trained on data from a completely different task (offline task). Normally, to do this, we would ideally want as large a volume of data from different subjects as possible. However, this is far from ideal as data collection is both time-consuming and expensive [1], further emphasising the importance of our approach. In our approach, we generate realistic data from a modest dataset containing data from nine subjects (detailed in Section 6.4.1). Different training dataset are used to train a SSVEP classifier model, which is able to accurately classify data from the three unseen subjects from the Online dataset (detailed in Section 6.4.1). The procedure of this evaluation is shown in Figure 6.8, where the SSVEP classifier model is trained using the Offline dataset and tested on the Online dataset.



(a) Evaluation method for AC-GAN.

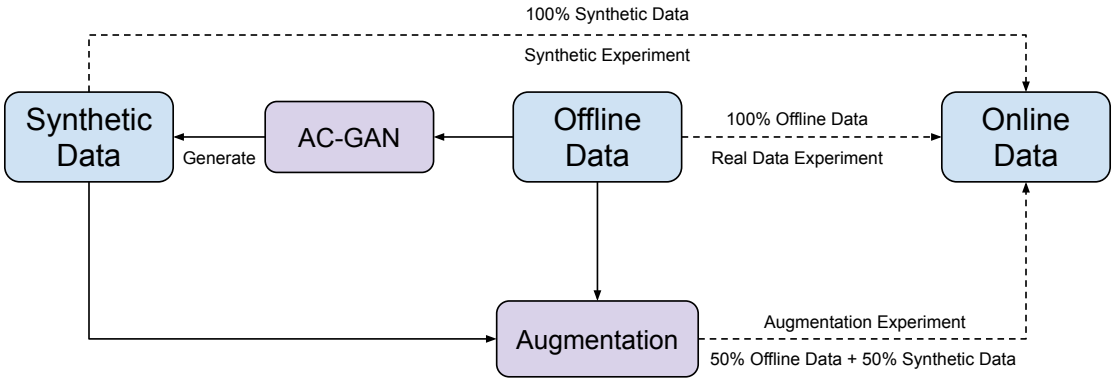


(b) Evaluation method for SIS-GAN.

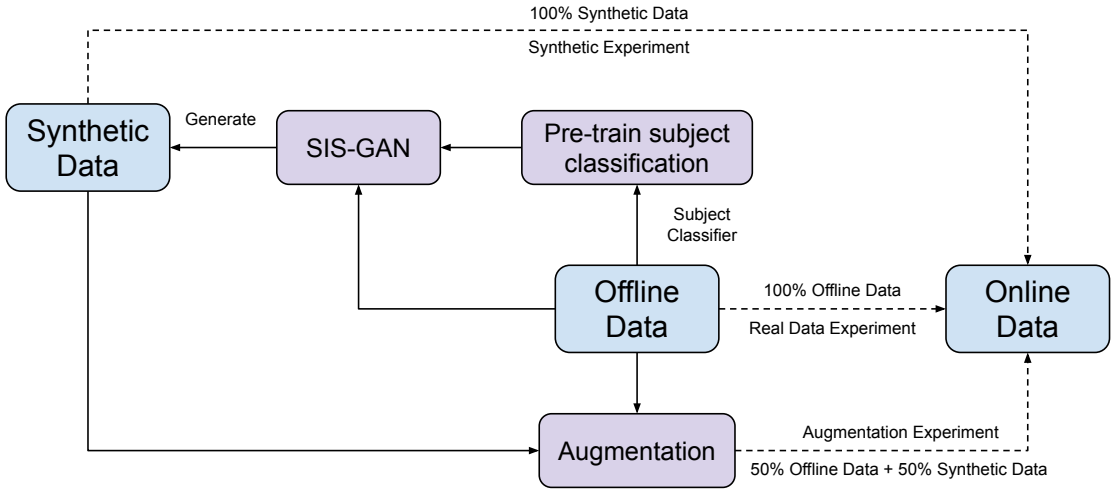
Figure 6.7: Evaluation method for SSVEP classification for leave-one-out unseen subject where the test data is from an unseen subject S0x with no prior training. Frozen subject-biometric classifier is used to train SIS-GAN. Dotted lines indicate SSVEP classification.

6.5 Experimental Results

The experimental results are presented here, with every result being the mean and standard deviation of ten different runs for each experiment, with a unique random seed used for every run. All experiments are performed using the same number of data points, with the augmented training dataset consisting of a mix of synthetic and real data at a ratio of 50:50. Final test experiments are performed on the real data.



(a) Evaluation method for AC-GAN.



(b) Evaluation method for SIS-GAN.

Figure 6.8: Evaluation method for SSVEP classification for cross-task unseen subject where the test data is from the online dataset with no prior training. Frozen subject-biometric classifier is used to train SIS-GAN. Dotted lines indicate SSVEP classification.

6.5.1 SSVEP Classification for a Single Subject

We have shown in our previous work in Chapter 5 that it is possible to generate realistic synthetic EEG signals containing SSVEP frequency features. The primary constraint in generating such synthetic EEG signals is the limited quantity of the real data available for training a model capable of generating synthetic data in the first place, as large datasets are required to train

accurate generative models. Since collecting EEG signals is time-consuming and expensive, we explore incorporating an auxiliary classifier into our approach for generating synthetic data as the auxiliary classification component essentially exposes the model to three times the amount of data provided to a vanilla approach such as DC-GAN.

	No augmentation	**With augmentation**	
Subject	**Real Data**	**DC-GAN**	**AC-GAN**
S01	0.727 ± 0.069	0.582 ± 0.089	**0.762 ± 0.036**
S02	0.843 ± 0.059	0.750 ± 0.089	**0.868 ± 0.020**
S03	**0.812 ± 0.048**	0.732 ± 0.058	0.780 ± 0.027
S04	0.783 ± 0.073	0.533 ± 0.087	**0.795 ± 0.027**
S05	0.807 ± 0.049	0.742 ± 0.076	**0.828 ± 0.023**
S06	0.948 ± 0.024	0.860 ± 0.080	**0.953 ± 0.016**
S07	0.962 ± 0.037	0.982 ± 0.029	**0.995 ± 0.008**
S08	**0.738 ± 0.086**	0.500 ± 0.113	0.708 ± 0.033
S09	0.863 ± 0.058	0.860 ± 0.032	**0.880 ± 0.025**
Mean	0.831 ± 0.098	0.727 ± 0.172	**0.841 ± 0.091**

Table 6.1: Mean accuracy with standard deviation when classifying SSVEP for unseen subject on offline dataset.

For this experiment, we attempt to classify SSVEP signals using data from a single subject – this means that a new model is trained for each of the nine subjects. The results from this experiment are presented in Table 6.1 and show that we are able to use synthetic data without losing much SSVEP classification performance. Additionally, the results show that there is a possibility that by augmenting synthetic data with real data, the classification result can actually be improved to some degree (from 83.1% for training on real data alone to 84.1% with data augmented by AC-GAN generated data). This can be seen as validation that realistic synthetic SSVEP data can indeed be generated. Based on the results presented here, and for the sake of brevity, for the remaining experiments we compare our results with AC-GAN as it has demonstrated the best performance on this task.

6.5.2 Subject-biometric Classification

In order to visualise the underlying problem in generalisation across subjects, we capture the embedding of the signals on the last layer of the SCU when training on nine subjects dataset to

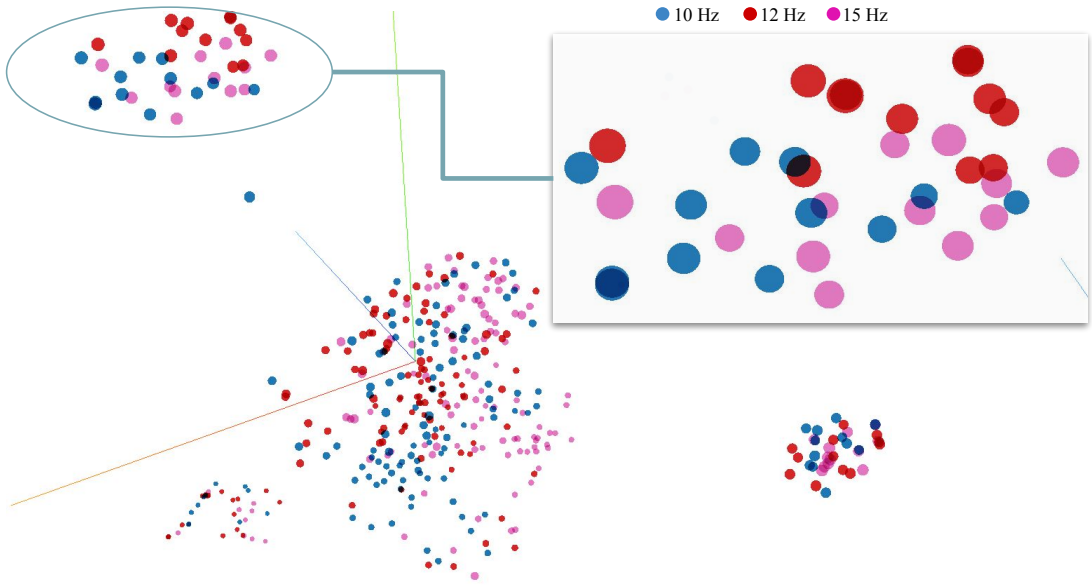
visualise the data points using T-distributed Stochastic Neighbor Embedding (t-SNE). Figure 6.9a shows the data points distribution with SSVEP frequencies as the ground truth label. The cluster of data points have mixed colour indicating the difficulty of the network to classify the signals of mixing subjects according to the frequency labels.

Furthermore, when the same data points are coloured according to the subject label (refer Figure 6.9b), the clusters of the same colour are much close together. This indicates the strong subject features contained in the signals even when the network is only provided with the frequency information - the distance of points from the same subjects are much closer than the distance of points from the same frequency.

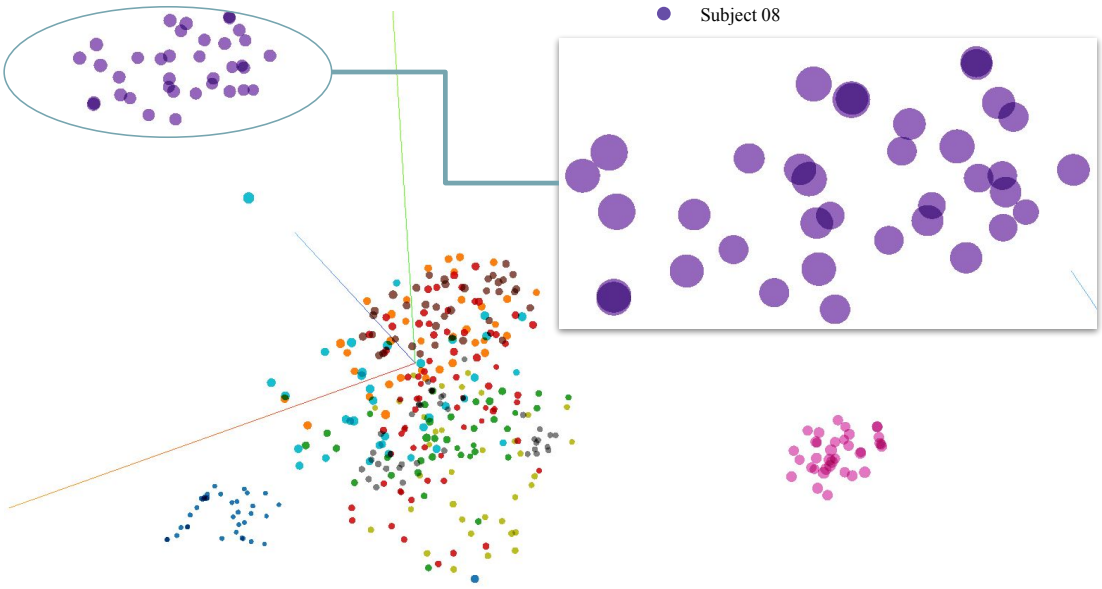
To further evaluate our hypothesis that SSVEP-based EEG signals can be classified in accordance with the subject-biometric, we train a convolutional model (similar to Chapter 3), where the subject is used as the training label, rather than the SSVEP frequency. The subject signals are gathered within the same session. A confusion matrix detailing the performance of the model on each of the nine subjects from the Offline dataset is seen in Figure 6.2. In addition, after using 10-fold cross validation, the model is able to produce a final mean test accuracy score over the nine subjects of 0.980 ± 0.050 when using the signals from the same session. This demonstrates that EEG signals contain within them enough unique information which is accurately able to identify the subject from which the data was recorded. This result also corroborates with other research showing subject information to have a large detrimental impact on model performance when performing other tasks [125, 98]. This is what primarily motivates the remainder of this work, as we try to eliminate these harmful subject-specific features when performing SSVEP classification.

6.5.3 SSVEP Classification for Unseen Subject : Leave-One-Out

The pattern of electrical voltage produced by the brain often differs from one subject to the next, resulting in difficulty for a model to correctly classify SSVEP signals from a subject whose signals is absent from a priori model training. To overcome such difficulty, we train a generative model to eliminate subject-specific features, potentially to be used on the unseen subject problem. Our proposed SIS-GAN (Section 6.3.3) introduces the subject invariant loss component based on the pre-trained frozen subject-biometric weight. Taking the probability values based on subject prediction on the synthetic data generated, this loss component minimises the maximum



(a) The distribution of data points embedding coloured by frequency.



(b) The distribution of data points embedding coloured by subject.

Figure 6.9: The distribution of data points embedding using T-distributed Stochastic Neighbor Embedding.

		Augmentation		Synthetic	
Unseen	**Real data**	**ACGAN**	**SIS-GAN**	**ACGAN**	**SIS-GAN**
S01	0.576 ± 0.055	0.534 ± 0.042	0.550 ± 0.028	0.617 ± 0.023	**0.665 ± 0.017**
S02	0.705 ± 0.009	0.711 ± 0.023	0.709 ± 0.032	0.619 ± 0.053	**0.714 ± 0.009**
S03	0.687 ± 0.016	0.688 ± 0.032	**0.732 ± 0.021**	0.604 ± 0.025	0.717 ± 0.006
S04	0.656 ± 0.010	0.581 ± 0.023	0.600 ± 0.028	0.627 ± 0.162	**0.667 ± 0.012**
S05	0.682 ± 0.027	0.694 ± 0.024	0.711 ± 0.026	0.380 ± 0.029	**0.768 ± 0.080**
S06	0.757 ± 0.045	0.710 ± 0.068	0.727 ± 0.032	0.627 ± 0.162	**0.815 ± 0.029**
S07	0.905 ± 0.005	0.940 ± 0.009	**0.943 ± 0.014**	0.588 ± 0.155	0.911 ± 0.025
S08	0.447 ± 0.046	0.426 ± 0.038	0.378 ± 0.046	0.343 ± 0.030	**0.485 ± 0.018**
S09	0.778 ± 0.012	0.807 ± 0.018	**0.814 ± 0.023**	0.710 ± 0.048	0.780 ± 0.009
Mean	0.688 ± 0.125	0.677 ± 0.146	0.685 ± 0.155	0.543 ± 0.150	**0.725 ± 0.113**

Table 6.2: Mean accuracy with standard deviation when classifying SSVEP for unseen subject on offline dataset.

probability forcing the generator to produce the synthetic EEG signals that are unable to classify subject with any reasonable accuracy.

For this experiment, we train both AC-GAN and SIS-GAN and evaluate the performance following the procedure outlined in Figure 6.7 in Section 6.4.2.

The results in Table 6.1 present how training on real data can be effective when the training and test data are from the same subject, with mean accuracy of 83.1% (Table 6.1 – Real Data), but when the settings of the experiment change to testing on previously unseen test subject, using real data alone for training is not effective, as the mean accuracy is reduced to 68.8% (Table 6.2 – Real Data). However, the results indicate how SIS-GAN can be a powerful tool to enhance unseen subject classification by generating subject-invariant synthetic data which can be used for training. As seen in Table 6.2, when the model is trained only on synthetic EEG signals generated by our SIS-GAN model and tested on real data, an accuracy level of up to accuracy 72.5% can be achieved, which is significantly higher than what a model trained on the same number of real data points can achieve (68.8%).

6.5.4 SSVEP Classification for Unseen Subject : Cross-Task

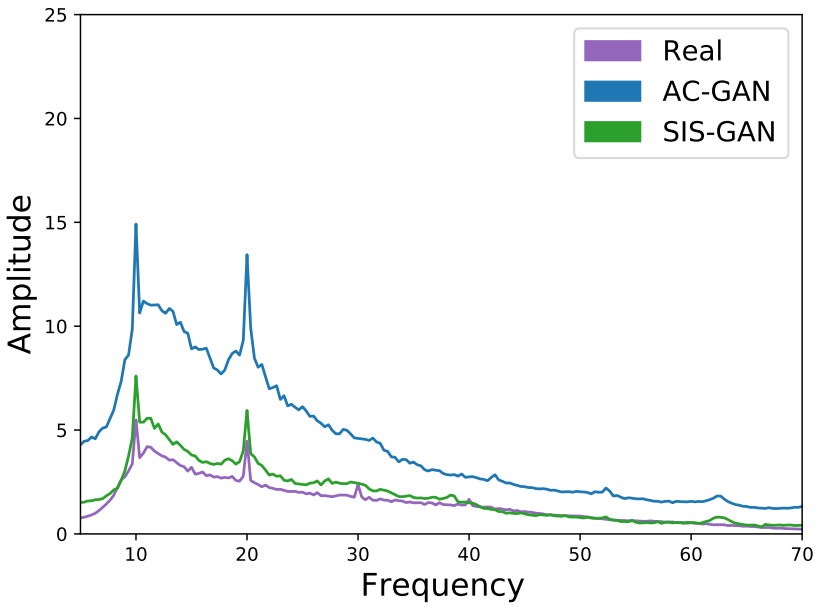
We demonstrate that our proposed generative model is able to produce realistic EEG signals containing SSVEP frequency information. To achieve this, we analyse a subset of the generated data, taken from a model trained on the Offline dataset, using a Fast-Fourier Transform (FFT). The FFT plots for the three SSVEP frequencies taken from SIS-GAN and AC-GAN, as well as the real data, are presented in Figure 6.10. The figure shows that when compared to the FFT taken from the real EEG signals, the generative models are capable of producing data which display the characteristic peak at the desired SSVEP frequency. It is interesting to note that although containing the correct peak, the data generated by AC-GAN often has a higher amplitude when compared to the real data.

To illustrate the potential of our generative models, we perform cross-task classification as in Section 6.4.2. Cross-task classification requires a model trained solely on a dataset focusing on one task to generalise well to another dataset focusing on another task. In our experimental setup, we evaluate the performance of a model trained on the Offline dataset when tested using the Online dataset. This is a common real-world problem as often BCI systems are trained beforehand on existing data and expected to perform on different subjects in real time on a range of tasks including teleoperation [16].

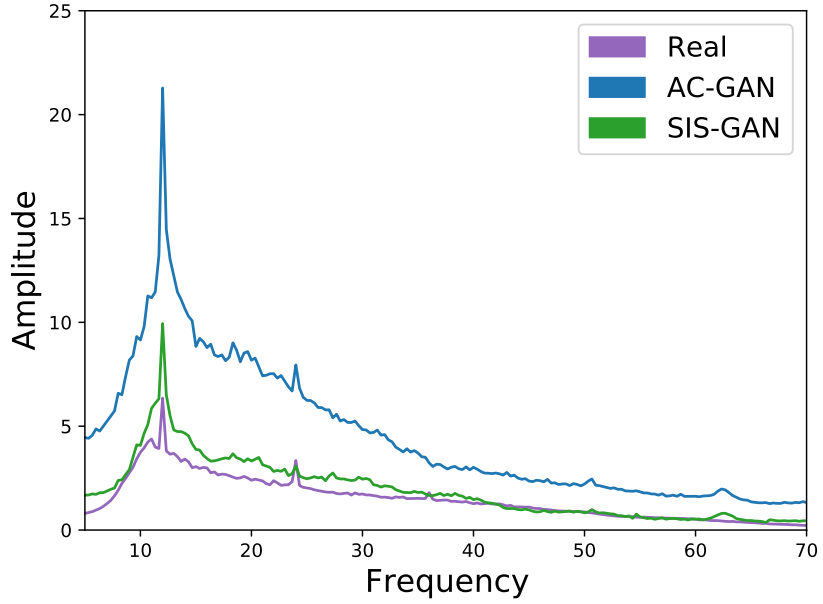
		Augmentation		**Synthetic**	
Subject	**Real data**	**ACGAN**	**SIS-GAN**	**ACGAN**	**SIS-GAN**
T01	0.736 ± 0.066	0.742 ± 0.071	0.756 ± 0.103	0.695 ± 0.100	**0.845 ± 0.019**
T02	0.508 ± 0.072	0.542 ± 0.062	0.592 ± 0.052	0.488 ± 0.083	**0.675 ± 0.016**
T03	0.240 ± 0.039	0.277 ± 0.067	0.303 ± 0.057	0.310 ± 0.060	**0.453 ± 0.017**
Mean	0.495 ± 0.211	0.520 ± 0.202	0.550 ± 0.201	0.498 ± 0.177	**0.660 ± 0.161**

Table 6.3: Mean accuracy with standard deviation for cross-task when classifying SSVEP on online subject dataset.

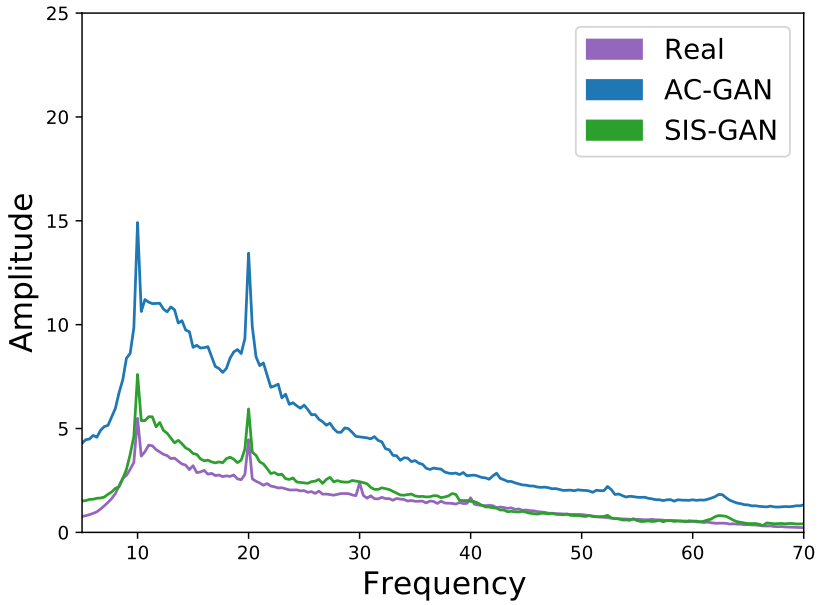
The experimental results are presented in Table 6.3. One of the most striking observations that can be made from the results is that almost all of the approaches improve upon just using the real data alone – bringing some evidence that using our synthetically generated data can improve cross-task performance. Perhaps the most interesting result is that training a classification model only with data generated using SIS-GAN can lead to the model outperforming a model trained



(a) SSVEP Signal at 10Hz.



(b) SSVEP Signal at 12Hz.



(c) SSVEP Signal at 15Hz.

Figure 6.10: Comparing Fast-Fourier Transforms (FFT) of real and synthetic data from all the generative models. The Synthetic data clearly displays the characteristic SSVEP frequency peaks and associated harmonics.

on the same number of real data points by over 16%. It is also of interest to observe that using the synthetic training data alone, SIS-GAN significantly outperforms AC-GAN, highlighting the importance and potential benefits associated with removing subject-specific features from EEG signals.

To further explore our approach, we investigate the properties of the subject-invariant synthetic data generated by SIS-GAN by visualising the softmax probability assigned to the generated data by the pre-trained subject-biometric classification network. Here, we hope to observe that the pre-trained network is unable to find any features in the generated data that can be used to classify the subject. The results from this experiment are displayed in Figure 6.11, where data generated by both the AC-GAN and SIS-GAN approaches are passed through the pre-trained subject network and the resulting softmax probability values are recorded. The figure shows that data generated by AC-GAN is often classified as belonging to either subject S03 or S04, meaning that strong subject-specific features are clearly still present in the data.

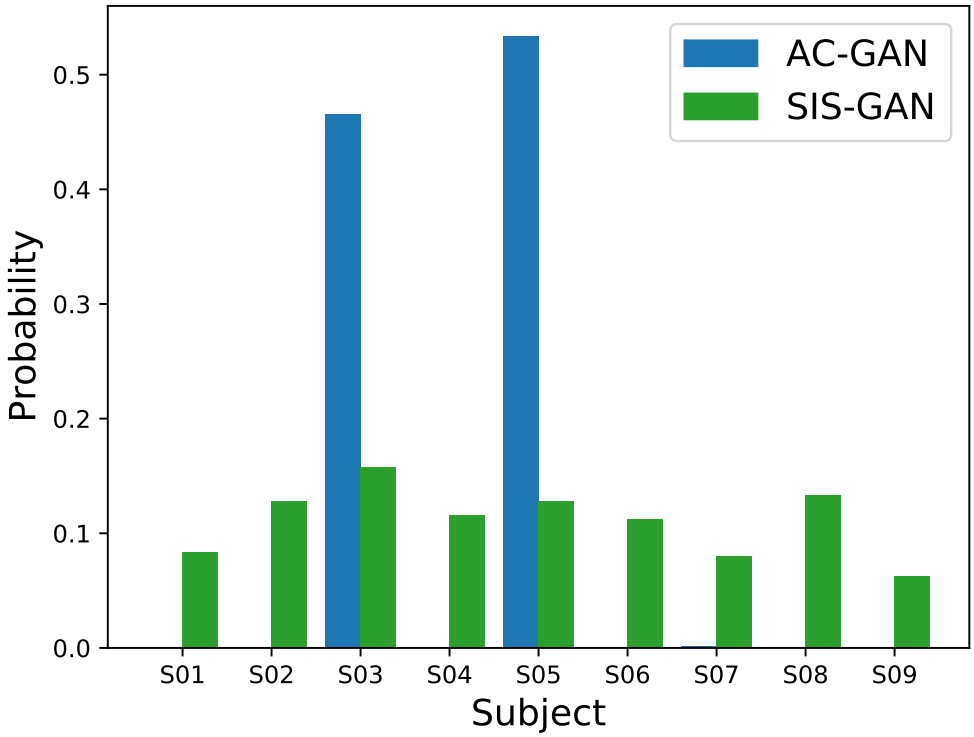


Figure 6.11: Softmax probability values taken from the pre-trained subject-biometric classification network for the generated data. A low consistent value distributed across all subject is best as it indicates the subject-biometric classification network is unable to find discriminative features.

However, the data generated by SIS-GAN has been assigned a low probability distributed across all nine subjects, indicating that the subject-biometric classification network is unable to find discriminative subject features in the data. This result further indicates that SIS-GAN produces subject invariant signals still containing the desired SSVEP frequency.

Additionally, the embedding of the subject invariant signals are captured using the same way as in Section 6.5.2 to visualise if the data points of this synthetic dataset are clustered according to the frequency classes. Figure 6.12 shows the data points are clustered according to the frequency.

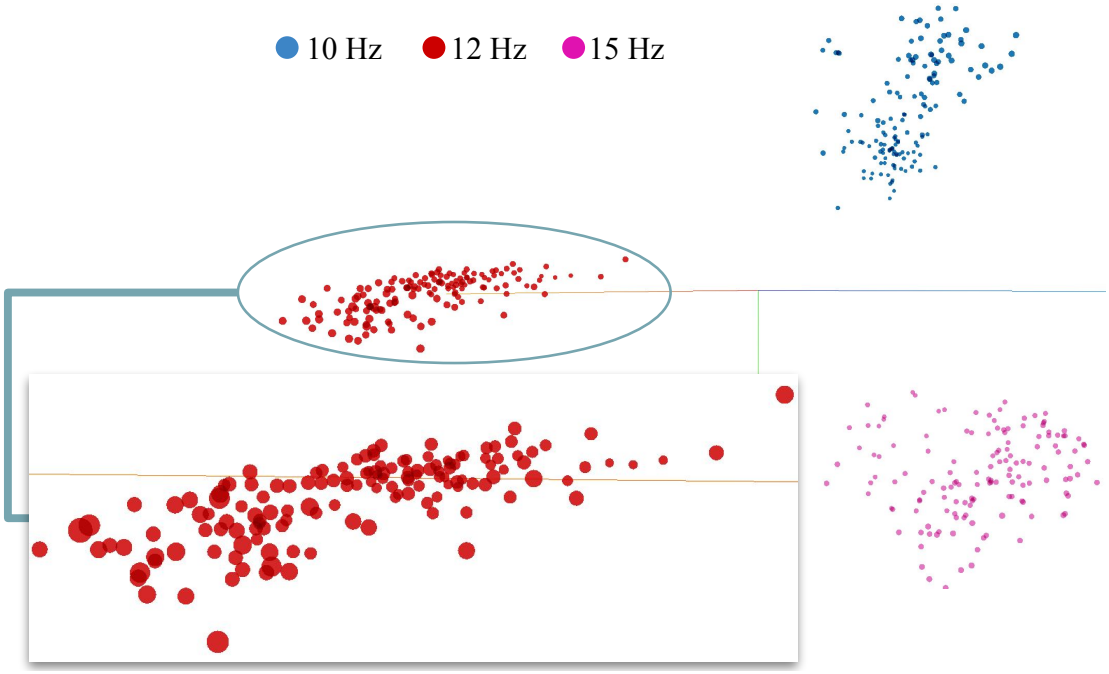


Figure 6.12: The distribution of data points embedding of subject invariant signals coloured by frequency.

6.6 Conclusion

Training learning-based BCI systems capable of generalising across subjects is currently an open research problem. In practice, many online BCI tasks require calibration to be performed for each new subject before an accurate prediction can be made. In this chapter, we have explored the possibility of generating new synthetic SSVEP-based dry-EEG data, via neural-based generative models, to help aid with this problem and move towards the removal of the calibration stage altogether. We have explored a novel approach, SIS-GAN to create subject-invariant data by attempting to remove subject-specific features, whilst preserving the SSVEP frequencies we are interested in. Earlier work discuss the difficulties often faced with the calibration stage and managed to tackle the session calibration within the same subject [112] and new class label [98], no other work, to date, has managed to classify completely unseen subjects without re-training, fine-tuning [125] and having a large dataset [98]. However, our experimental results demonstrate the efficacy of the synthetically-generated data in improving the performance of a downstream

SSVEP frequency classification model. In fact, by training only on synthetic data, we are able to improve the unseen subject generalisation when performing zero-calibration classification by up to 16% for cross-task classification.

Whilst this work proves the potential of a subject invariant generative model to tackle the usually required calibration stage, there is a lack of evaluation performed on a real-time BCI application to further show the potential of removing the calibration stage completely. Apart from that, neural-based models are known for having a large computational requirement. Although our proposed subject invariant GAN improves classification performance, it requires two additional networks (auxiliary and subjection classification modules) on top of the normal GAN architecture – thus increasing the computational requirement.

For future work, incorporating this zero-calibration BCI application into a real-time teleoperation system can provide excellent testing opportunities. Additionally, the subject invariance within the synthetic data can be further enhanced via a more targeted objective function utilising something analogous to the inception score specifically aimed at reducing subject-specific features within the data.

Chapter 7

Conclusion

Despite the continuously growing enthusiasm in the research of Brain-Computer Interface, especially in the field of Electroencephalography-based applications, there is a distinct lack of applicability outside of the research laboratory [30, 160]. This is due to some fundamental constraints such as the rigid experimentation protocol, limitations within the technology and the challenges of accurate signal decoding, especially in making a system work in real-time. This thesis focuses on exploring the use of recent advances in deep neural networks, to tackle certain critical limitations of SSVEP-based BCI applications, such as rigid experimentation procedure, limited frequency ranges, hand-crafted classifier, small dataset and expensive calibration stage.

With recent technology, there are possible alternatives to eliminate the typical cumbersome experimental protocol and make it easy to be used frequently, even in a real-time application. Often the drawback of such an alternative solution is the poorer quality of the bio-signals collected, resulting in a substantially more challenging signal decoding and classification task – thus possible in producing inaccurate translations [154]. The recent advances in the Machine Learning field, notably in deep neural networks that have the ability to discriminate data based on characteristic features, is a possible way to tackle the fundamental constraints within the EEG-based BCI application. The ability of DNN to autonomously learn the best representation, even from potentially noisy input signals, make it possible to ease down the rigid application procedure. Additionally, taking advantage of the accomplishment of DNN in the object detection task, it is possible for SSVEP stimuli to be implemented in such way a that they appear more natural

to subjects by exploiting the objects detected within the scene as stimulus for the application. Finally, the versatility of DNN to generate entirely new and realistic data points, which match the distribution of a given target dataset, draws the possibility for the algorithm to generate sets of synthetic data that potentially address the issue of typically small datasets and an expensive calibration stage [82].

In this thesis, we have explored the potential solutions using DNN for tackling some of the fundamental constraints within BCI applications, especially in the applicability of EEG-based SSVEP applications. Three potential techniques using DNN are bio-signal pattern recognition, object detection in video streams and synthetic bio-signal generation. We established the primary research aim and objectives of the thesis in Chapter 1 where the primary objectives are to provide less rigid BCI experimental setup, to design a more natural SSVEP stimuli and to eliminate the typical calibration stage.

7.1 Review of Key Contributions

This thesis presents various contributions to the state-of-the-art within the area of EEG-based BCI. In this section, we present a summary of the contributions including qualitative evaluation and analysis of the research contributions made during this thesis.

In Chapter 3, we design a more applicable SSVEP-based system that eliminates rigid procedure and reduces the preparation time for BCI experiments. Unlike previous studies [115, 226, 120], throughout this study, we use the state-of-the-art dry-EEG headset for the EEG signals data collection. This proposed system has successfully provided a BCI application with a flexible platform by using open source programming and using a commercial dry-sensor technology where the experiment can be done without a specific physical shield room. This is significant as often within the field of BCI, most experiments are rigid, hard to replicate and there can be software compatibility limitations. The focus of the chapter is to demonstrate the ability of such a system to produce correct prediction of signals. The experiment performed to collect the bio-signals is the most crucial part in BCI, as often less rigid procedure leads to more noisy signals that are hard to decode. Although it is good to be able to make the procedure more flexible, it is equally important to ensure that the signals collected can be decoded accurately. This thesis demonstrates the ability of CNN to decode such noisier signals. Instead of using

traditional machine learning as in [133, 31, 135], here, we propose SSVEP Convolutional Unit to eliminate the traditional hand-crafted manual feature extraction to perform the classification of raw dry-EEG SSVEP. Comparing to other DNN and the traditional ML approaches, our SCU model demonstrates the ability to classify the dry-EEG signals with greater levels of accuracy, importantly including generalisation across subjects and even to an entirely unseen subject. This shows the ability of this less rigid experimentation to replace the traditional system and becomes the basis of our studies. This procedure is used as the framework for the remainder of the work presented in this thesis, which will focus on using it to make systems applicable for daily use in the real world.

The ultimate aim of BCI is to provide a communication medium for patients suffering from conditions like LIS. Therefore, a successful BCI system is one that is able to be used under real-time operating conditions. In Chapter 4, we have successfully integrated the experimentation procedure of collecting dry EEG signals to execute a BCI application in real-time, where the BCI subjects can communicate to an external device (a humanoid robot). We designed a novel variable position and size SSVEP BCI stimuli for the navigation of a humanoid robot, that appears more natural for subjects. Within the live video stream from a teleoperated humanoid robot traversing a real-world environment, we employ the use of a high-performance scene object detection model to identify objects in real-time to replace the traditional SSVEP stimuli. Traditional SSVEP stimuli can be fatiguing for users, especially when viewing static stimuli for extended periods. Some prior studies did not integrate the stimuli on the robot video feedback [257, 64, 137], making it harder for the subject to concentrate on the stimuli whilst also viewing the feedback display. Unlike the prior literature [58, 257, 64, 137], our variable stimuli are varied in terms of pixel pattern, size and on-screen position, in-conjunction with the changing nature of the environment the robot is navigating through. We demonstrate the capability of the proposed stimuli to be used in BCI teleoperation application in real-time through a natural indoor environment by evaluating the accuracy and the ITR. Our results demonstrate a mean accuracy of 0.85 across all subjects which is comparable to [258] which obtained 0.88 accuracy, despite our work using a variable SSVEP stimuli.

In Chapter 5, we explore a variety of unsupervised models such as a Generative Adversarial Network, a Wasserstein Generative Adversarial Network and a Variational Auto-Encoder, to generate synthetic EEG bio-signals containing SSVEP signals. There are numerous shortcomings and limitations within the current state-of-the-art that lead to the low reliability and usability of BCI, such as the requirement for careful per-subject and per-session calibration that leads to

limited quantities of data as collecting bio-signal can be expensive and time-consuming [140, 142]. We tackle the problem of reliable data shortages by generating meaningful and realistic synthetic EEG signals potentially capable of improving the SSVEP classification models trained on EEG signals. To the best of our knowledge at the time when this work was undertaken, this is the first work in the literature to explore the use of neural-based generative models to create dry-EEG signals containing SSVEP information. We demonstrate the efficacy of the realistic synthetic EEG signals using both visualisation and accuracy evaluation. By using pre-training of the classifier and through dataset augmentation, the use of synthetic EEG signals can improve the classification of real-world EEG signals and can increase the convergence rate of classification models, thus resulting in the observation that smaller quantities of real-world training data are required. In comparison to the prior work, we generate synthetic EEG signals containing SSVEP signals [89, 53] directly via end-to-end training without first transforming the signals into different domains [253, 191].

Finally, in Chapter 6, we implement an improved EEG generative model by taking into consideration the biased subject features present within the bio-signals. EEG signals have been known to be highly subject and session dependent, containing a distinctive subject bias through subject-biometric features [178]. This can potentially result in the need for a calibration stage at the beginning of an experiment which becomes a hindrance for a practical real-world application [112, 142]. We propose a single network, SSVEP Subject Invariant GAN (SIS-GAN) that can produce realistic synthetic EEG signals without containing subject-specific features. The synthetically-generated subject invariant EEG signals are not only visually realistic, but also offer superior performance for downstream signal decoding tasks. We evaluate the SSVEP classification performance on unseen subjects within the same task. This shows a performance improvement compared to using real-data alone. Additionally, downstream models trained on data generated by our SIS-GAN approach are capable of SSVEP classification of unseen subjects under the setting of a different task (cross-task) where the results are shown to perform well when classifying online signals having only seen offline signals during training. This is a promising achievement and is a step towards real-time BCI applications with no mandatory calibration stage even on totally new subject [112, 98] and without complex re-training model [125].

7.2 Review of Research Aim and Objectives

Throughout the study undertaken within this thesis, we have successfully accomplished the primary aim of this research and the associated objectives:

- Firstly, a more flexible experimentation procedure has been proposed in Chapter 3 with the usage of commercial dry-sensor technology. This procedure is not only eliminating the cumbersome pre and post-procedure related to the gel application but other limiting factors like the experimental data acquisition interval, the need for a physical shield, the bulky size of the headset and the wired connectivity. Through the use of a DNN, this less rigid procedure is still able to provide an accurate signal prediction for the BCI application.
- Secondly, by incorporating state-of-the-art computer vision techniques for object detection, we design a novel variable SSVEP interface as in Chapter 4 for real-time teleoperation BCI application. These stimuli are based on the object detected within the environment the humanoid robot is navigating through. The ability of this complete application is accessed by the accuracy performance and the information transfer rate of the real-time navigation.
- Finally, a variety of synthetic EEG-data generation models have been devised and evaluated to tackle the session and subject variant problems in the BCI field. We undertake the session variant problems by generating synthetic EEG signals containing SSVEP signals in Chapter 5 that not only improve the classification accuracy but also to provide a measure for the complexity in collecting a large EEG dataset. In Chapter 6, a novel Generative Adversarial Network, SIS-GAN has been proposed, resulting in synthetically-generated subject-invariant EEG signals that capable of accurate SSVEP classification for unseen subjects, suitable for real-time BCI applications with near-zero calibration.

7.3 Limitations

Although this thesis has tackled some of the constraints within the area of SSVEP-based dry-EEG BCI, there are some limitations within the scope of this thesis:

- *Small dataset* - The results presented in this thesis are based on three different experiments comprising signals taken from a total of 12 subjects. Whilst the results were positive, a larger and more varied dataset should be used to create further justification of the overall contributions especially proving the ability of our proposed system to be used across different subjects and tasks. This size of dataset is inherently link to the complexity of collecting BCI subject data and is typical of prior work in the field [253, 145, 53].

- *Supervised Experiment* - The real-time evaluation presented in Chapter 4 address the limitations from Chapter 3 in term of stimuli and real-time execution, but due to the limitation of the on-board camera on our humanoid robot that only provides a 2D image, the experiments have been conducted in a supervised manner. The pre-prepared floor plan is used for the robot navigation as it is important to prove the variable stimuli concept where essentially the distance and the direction from the robot to an object is based on the object height. Additionally, the experiment uses a default condition where the humanoid robot will always move towards the ground truth label. Any incorrect predictions, whilst being incorporated in the reported accuracy score, do not contribute towards improving the underlying SSVEP classification model. Changes could be made to this system to try and incorporate an online-learning [94] style of model training.

- *Large Computational Requirements* - Neural-based models are known for having a large computational usage and using two networks competing with one another (Discriminator and Generator) requires even greater computational resources. Although our proposed subject invariant GAN improves the non-calibrated classification, it has the requirement of the addition auxiliary and subject classifier on top of the already complex GAN structure – meaning a even greater computational demand [127].

7.4 Future Work

Based on the limitations explored in the previous section (Section 7.3, in this section we will focus on the work that could be undertaken to improve the current study and further aspects which could be performed as an expansion to the current study.

7.4.1 Improvements to Current Work

Despite the successes of this thesis in tackling the constraints within a BCI application, explored in this thesis, there are some areas in which the research could have been improved. This is mostly laid out in the conclusion sections for the individual chapters, however, the major items are summarised in this section.

Throughout this thesis we presented results on a modest dataset involving 12 subjects across different tasks, future work could involve the collection of a larger datasets taken from a broader set of subjects [173, 68, 234]. Additionally, it is worth considering aspects of robot interaction within the environment, for example, having obstacles for it to avoid [96] or having the robot freely roaming around the scene, or even perhaps conducting experiments involving an outdoor scene [64]. This could help in making the application system more robust and flexible. Finally, performing a real-time experiment with data taken from the Subject Invariant SSVEP GAN could provide excellent testing opportunities to established if a zero-calibration BCI application is truly possible using a real patient.

7.4.2 Expansions to the Work

Based on the work undertaken in this thesis, there are several aspects that could be expanded for future improvement to the BCI system, which are outlined below.

Domain Adaptation: Although there are some available open-source dataset on EEG signals, to our knowledge, there are only three SSVEP-based dataset available [173, 68, 234]. Even within these SSVEP dataset, often the frequency classes used and the application being performed are different. Potential future work could look into domain adaptation by pre-training a network on a larger dataset from a different domain and transferring the knowledge to our dataset. Cross-dataset domain adaptation has been performed within EEG-based emotion recognition work as in [117]. The authors used two open-source datasets on emotion recognition collected using different EEG devices and on different emotion tasks. By using five different domain adaptation methods, there is some improvement in the classification performance across the dataset.

Creation of Benchmark Dataset: As previously mentioned, there are only a small number of publicly available datasets on SSVEP-based EEG signals, making it challenging for use with

machine learning. Especially in the computer vision field, large datasets have helped the classifier to learn better when the dataset is comprehensive, covering a broader range of diverse and unique data [57]. The well-known example is the large dataset of Imagenet [57] that has became one of the fundamental datasets in computer vision. It has allowed for direct comparison between approaches without methods having to be reimplemented [202]. The creation of a benchmark dry-EEG based SSVEP dataset could potentially provide a better starting point for approaches to try and solve the remaining BCI constraints.

Depth Variable Stimuli: Instead of using a 2D image camera as in Chapter 4, as we suggested for further improvement of this work, it may be possible to use a depth camera where the distance and angle of the objects can be determined from the additional depth information [12]. By such ability, it would be interesting to use the different depth of the objects as the label as the flickering frequency instead of using the object type itself. It may give the freedom to navigate the humanoid robot according to the distance of the object.

Real Patient Evaluation: To further prove the applicability of our proposed work, the improved BCI system would benefit from being tested on real patients. To do so, there are important aspects to be taken into consideration before experimentation. For example, if the EEG signals may be weaker for the some patients as compared to healthy subjects [157], if the patients have epilepsy [103] or remission of epilepsy [77] and the possible drop in the performance [124].

Interpretable Models for EEG: Neural models are known to be black-box, where understanding the reasoning behind the decision taken is challenging [60]. Most of machine learning based BCI applications cannot be interpreted based on what the model learns from the signals [144]. On top of that, the human brain is deeply complex and hard to understand. This manifests in the requirement of calibration being needed to make the system more accurate, as sometimes within differing sessions with the same subject, varying signal patterns are present – making even cross-session generalisation a challenge. Interpretable DNN is a progressing field, as an increasing number of studies are attempting to understand the meaning behind the decision made by a particular model [106, 37]. There are some initial works focusing on how to interpret the EEG signals, either pre-classification by representing the EEG sleep signals in time-frequency images as the input to CNN [231], or post-classification by producing heatmaps to see the relevance of the EEG data points based on the classification [221].

"We took to the skies in the name of discovery,
to find something new, to change the world.
But you don't change the world simply by looking at it.
You change it through the way you choose to live in it."
- Amelia Wren, The Aeronouts 2019

Bibliography

[1] Sherif M Abdelfattah, Ghodai M Abdelrahman, and Min Wang. Augmenting The Size of EEG datasets Using Generative Adversarial Networks. In *International Joint Conference on Neural Networks*, pages 1–6. IEEE, 2018.

[2] Sarah N Abdulkader, Ayman Atia, and Mostafa-Sami M Mostafa. Brain Computer Interfacing: Applications and Challenges. *Egyptian Informatics Journal*, 16(2):213–230, 2015.

[3] Reza Abiri, Soheil Borhani, Eric W Sellers, Yang Jiang, and Xiaopeng Zhao. A Comprehensive Review of EEG-based Brain-Computer Interface Paradigms. *Journal of Neural Engineering*, 16(1):011001, 2019.

[4] Norhafizan Ahmad, Raja Ariffin Raja Ghazilla, and Muhammad Zikril Hakim MD Azizi. Steady State Visual Evoked Potential based BCI as Control Method for Exoskeleton: A Review. *Malaysian Journal of Public Health Medicine*, 16(Sppl. 1):86–94, 2016.

[5] Nasim Alamdari, Ali Haider, Riadh Arefin, Ajay K Verma, Kouhyar Tavakolian, and Reza Fazel-Rezai. A Review of Methods and Applications of Brain Computer Interface Systems. In *International Conference on Electro Information Technology*, pages 0345–0350. IEEE, 2016.

[6] X An, D P Kuang, X J Guo, Y L Zhao, and L H He. A Deep Learning Method for Classification of EEG Data Based on Motor Imagery. *Intelligent Computing in Bioinformatics*, 8590:203–210, 2014.

[7] Søren K Andersen and Matthias M Müller. Driving Steady-State Visual Evoked Potentials at Arbitrary Frequencies using Temporal Interpolation of Stimulus Presentation. *BMC Neuroscience*, 16(1):95, 2015.

[8] Martin Arjovsky and Léon Bottou. Towards Principles for Training Generative Adversarial Networks. In *International Conference Learning Representations*, pages 1–17, 2017.

[9] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein Generative Adversarial Networks. In *International Conference Machine Learning*, pages 214–223, 2017.

[10] Pablo Arnau-González, Miguel Arevalillo-Herráez, Stamos Katsigiannis, and Naeem Ramzan. On The Influence of Affect in EEG-based Subject Identification. *IEEE Transactions on Affective Computing*, 2018.

[11] Pablo Arnau-Gonzalez, Stamos Katsigiannis, Naeem Ramzan, Debbie Tolson, and Miguel Arevalillo-Herrez. ES1D: A Deep Network for EEG-based Subject Identification. In *International Conference on Bioinformatics and Bioengineering*, pages 81–85. IEEE, 2017.

[12] Amir Atapour-Abarghouei and Toby P Breckon. Veritatem Dies Aperit-temporally Consistent Depth Prediction Enabled by a Multi-task Geometric and Semantic Scene Understanding Approach. In *Computer Vision and Pattern Recognition*, pages 3373–3384. IEEE, 2019.

[13] Nik Khadijah Nik Aznan, Amir Atapour-Abarghouei, Stephen Bonner, Jason D Connolly, Noura Al Moubayed, and Toby P Breckon. Simulating Brain Signals: Creating Synthetic EEG Data via Neural-based Generative Models for Improved SSVEP Classification. In *International Joint Conference on Neural Networks*, pages 1–8. IEEE, 2019.

[14] Nik Khadijah Nik Aznan, Amir Atapour-Abarghouei, Stephen Bonner, Jason D Connolly, and Toby P Breckon. Leveraging Synthetic Subject Invariant EEG Signals for Zero Calibration BCI. In *International Conference on Pattern Recognition*. IEEE, 2020.

[15] Nik Khadijah Nik Aznan, Stephen Bonner, Jason Connolly, Noura Al Moubayed, and Toby P Breckon. On the Classification of SSVEP-based dry-EEG Signals via Convolutional Neural Networks. In *International Conference on Systems, Man, and Cybernetics*, pages 3726–3731. IEEE, 2018.

[16] Nik Khadijah Nik Aznan, Jason D Connolly, Noura Al Moubayed, and Toby P Breckon. Using Variable Natural Environment Brain-Computer Interface Stimuli for Real-Time Humanoid Robot Navigation. In *International Conference Robotics and Automation*, pages 4889–4895. IEEE, 2019.

[17] Nik Khadijah Nik Aznan, Kyung-Moo Huh, and Yeon-Mo Yang. EEG-based Motor Imagery Classification in BCI System by Using Unscented Kalman Filter. *International Journal of Information and Communication Technology*, 9(4):492–508, 2016.

[18] Shaojie Bai, J Zico Kolter, and Vladlen Koltun. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv preprint arXiv:1803.01271*, 2018.

[19] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass Brain–Computer Interface Classification by Riemannian Geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2011.

[20] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Classification of Covariance Matrices using a Riemannian-based Kernel for BCI Applications. *Neurocomputing*, 112:172–178, 2013.

[21] D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.

[22] BBCI. BCI Competition IV. 2009.

[23] Jonathan Becedas. Brain-Machine Interfaces: Basis and Advances. *IEEE Systems, Man and Cybernetics*, 42(6):825–836, 2012.

[24] Mihaly Benda, Piotr Stawicki, Felix Gembler, Roland Grichnik, Aya Rezeika, Abdul Saboor, and Ivan Volosyak. Different Feedback Methods for an SSVEP-based BCI. In *International Conference of the Engineering in Medicine and Biology Society*, pages 1939–1943. IEEE, 2018.

[25] Vitoantonio Bevilacqua, Giacomo Tattoli, Domenico Buongiorno, Claudio Loconsole, Daniele Leonardis, Michele Barsotti, Antonio Frisoli, and Massimo Bergamasco. A Novel BCI-SSVEP based Approach for Control of Walking in Virtual Environment using a Convolutional Neural Network. In *International Joint Conference on Neural Networks*, pages 4121–4128. IEEE, 2014.

[26] Luzheng Bi, Xin-An Fan, and Yili Liu. EEG-based Brain-Controlled Mobile Robots: A Survey. *IEEE Transactions on Human-Machine Systems*, 43(2):161–176, 2013.

[27] Stephen Bonner, Amir Atapour-Abarghouei, Philip T Jackson, John Brennan, Ibad Kureshi, Georgios Theodoropoulos, Andrew Stephen McGough, and Boguslaw Obara. Temporal Neighbourhood Aggregation: Predicting Future Links in Temporal Graphs via

Recurrent Variational Graph Convolutions. In *International Conference on Big Data*, pages 5336–5345. IEEE, 2019.

[28] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library.* O'Reilly Media, Inc., 2008.

[29] Szilárd Bulárka and Aurel Gontean. EEG Pattern Recognition Techniques Review. In *International Symposium for Design and Technology in Electronic Packaging*, pages 273–276. IEEE, 2015.

[30] Szilárd Bulárka and Aurel Gontean. Brain-Computer Interface Review. In *International Symposium on Electronics and Telecommunications*, pages 219–222. IEEE, 2016.

[31] Daniel E. Callan, Gautier Durantin, and Cengiz Terzibas. Classification of Single-trial Auditory Events using Dry-wireless EEG during Real and Motion Simulated Flight. *Frontiers in Systems Neuroscience*, 9:1–12, 2015.

[32] Lei Cao, Tianyu Liu, Lusong Hou, Zijian Wang, Chunjiang Fan, Jie Li, and Haoran Wang. A Novel Real-Time Multi-Phase BCI Speller Based on Sliding Control Paradigm of SSVEP. *IEEE Access*, 7:133974–133981, 2019.

[33] Sarah N Carvalho, Thiago BS Costa, Luisa FS Uribe, Diogo C Soriano, Glauco FG Yared, Luis C Coradine, and Romis Attux. Comparative Analysis of Strategies for Feature Extraction and Classification in SSVEP BCIs. *Biomedical Signal Processing and Control*, 21:34–42, 2015.

[34] Schubert R Carvalho, Iraquitan Cordeiro Filho, Damares Oliveira De Resende, Ana Carolina Siravenha, Cleidson De Souza, Henrique Galvan Debarba, Bruno Gomes, and Ronan Boulic. A Deep Learning Approach for Classification of Reaching Targets from EEG Images. In *SIBGRAPI Conference on Graphics, Patterns and Images*, pages 178–184. IEEE, 2017.

[35] Yongwook Chae, Jaeseung Jeong, and Sungho Jo. Toward brain-actuated humanoid robots: Asynchronous direct control using an EEG-Based BCI. *IEEE Transactions on Robotics*, 28(5):1131–1144, 2012.

[36] Elisavet Chatzilari, Georgios Liarios, Kostas Georgiadis, Spiros Nikolopoulos, and Yiannis Kompatsiaris. Combining the Benefits of CCA and SVMs for SSVEP-based BCIs in Real-world Conditions. In *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care*, pages 3–10. ACM, 2017.

[37] Chaofan Chen, Oscar Li, Daniel Tao, Alina Barnett, Cynthia Rudin, and Jonathan K Su. This Looks Like That: Deep Learning for Interpretable Image Recognition. In *Advances in Neural Information Processing Systems*, pages 8930–8941, 2019.

[38] Shih-Chung Chen, Yeou-Jiunn Chen, Ilham AE Zaeni, and Chung-Min Wu. A Single-Channel SSVEP-Based BCI with a Fuzzy Feature Threshold Algorithm in a Maze Game. *International Journal of Fuzzy Systems*, 19(2):553–565, 2017.

[39] Xiaogang Chen, Yijun Wang, Shangkai Gao, TzyyPing Jung, and Xiaorong Gao. Filter Bank Canonical Correlation Analysis for Implementing a High-speed SSVEP-based Brain–Computer Interface. *Journal of Neural Engineering*, 12(4):46008, 2015.

[40] Xiaogang Chen, Yijun Wang, Masaki Nakanishi, Xiaorong Gao, Tzyy-Ping Jung, and Shangkai Gao. High-Speed Spelling with a Noninvasive Brain–computer Interface. *Proceedings of the National Academy of Sciences*, 112(44):E6058–E6067, 2015.

[41] Xiaogang Chen, Bing Zhao, and Xiaorong Gao. Noninvasive Brain-computer Interface Based High-level Control of a Robotic Arm for Pick and Place Tasks. In *International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, pages 1193–1197. IEEE, 2018.

[42] Xiaogang Chen, Bing Zhao, Yijun Wang, and Xiaorong Gao. Combination of High-frequency SSVEP-based BCI and Computer Vision for Controlling a Robotic Arm. *Journal of neural engineering*, 2018.

[43] Sylvain Chevallier, Emmanuel Kalunga, Quentin Barthélemy, and Florian Yger. Riemannian Classification for SSVEP based BCI: Offline versus Online Implementations. In *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances*. 2018.

[44] Yu M Chi, Yijun Wang, Yu-Te Wang, Tzyy-Ping Jung, Trevor Kerth, and Yuchen Cao. A Practical Mobile Dry EEG System for Human Computer Interfaces. In *International Conference on Augmented Cognition*, pages 649–655. Springer, 2013.

[45] Ching-Yu Chiu, Avinash K Singh, Yu-Kai Wang, Jung-Tai King, and Chin-Teng Lin. A Wireless Steady State Visually Evoked Potential-based BCI Eating Assistive System. In *International Joint Conference on Neural Networks*, pages 3003–3007. IEEE, 2017.

[46] Andrei Chiuzbaian, Jakob Jakobsen, and Sadasivan Puthusserypady. Mind Controlled Drone: An Innovative Multiclass SSVEP based Brain Computer Interface. In *International Winter Conference on Brain-Computer Interface*, pages 1–5. IEEE, 2019.

[47] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations using RNN Encoder-decoder for Statistical Machine Translation. *arXiv preprint arXiv:1406.1078*, 2014.

[48] Bongjae Choi and Sungho Jo. A Low-Cost EEG System-Based Hybrid Brain-Computer Interface for Humanoid Robot Navigation and Recognition. *PLoS ONE*, 8(9), 2013.

[49] Bongjae Choi and Sungho Jo. Unification of Neural Systems between Human and Humanoid Robot. In *Asian Control Conference*, pages 1–6. IEEE, 2015.

[50] Francois Chollet. *Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*. MITP-Verlags GmbH & Co. KG, 2018.

[51] Cognionics. Quick-20 Dry EEG Headset. 2017.

[52] Adrien Combaz, Camille Chatelle, Arne Robben, Gertie Vanhoof, Ann Goeleven, Vincent Thijs, Marc M Van Hulle, and Steven Laureys. A Comparison of Two Spelling Brain-computer Interfaces based on Visual P3 and SSVEP in Locked-in Syndrome. *PloS one*, 8(9):e73691, 2013.

[53] Isaac A Corley and Yufei Huang. Deep EEG super-resolution: Upsampling EEG spatial resolution with Generative Adversarial Networks. In *International Conference Biomedical & Health Informatics*, pages 100–103. IEEE, 2018.

[54] Saumitro Dasgupta, Michael Fanton, Jonathan Pham, Michael Willard, H Nezamfar, B Shafai, and D Erdogmus. Brain Controlled Robotic Platform using Steady State Visual Evoked Potentials Acquired by EEG. In *Conference on Signals, Systems and Computers*, pages 1371–1374. IEEE, 2010.

[55] W David Hairston, Keith W Whitaker, Anthony J Ries, Jean M Vettel, J Cortney Bradford, Scott E Kerick, and Kaleb McDowell. Usability of Four Commercially-oriented EEG Systems. *Journal of Neural Engineering*, 11(4):046018, 2014.

[56] Omid Dehzangi and Muhamed Farooq. Portable Brain-Computer Interface for the Intensive Care Unit Patient Communication Using Subject-Dependent SSVEP Identification. *BioMed research international*, 2018, 2018.

[57] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale Hierarchical Image Database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. IEEE, 2009.

[58] Pablo F Diez, Vicente A Mut, Eric Laciar, and Enrique M Avila Perona. Mobile Robot Navigation with a Self-paced Brain–computer Interface based on High-frequency SSVEP. *Robotica*, 32(5):695–709, 2014.

[59] James Dowsett, Marianne Dieterich, and Paul CJ Taylor. Mobile Steady-state Evoked Potential Recording: Dissociable Neural Effects of Real-world Navigation and Visual Stimulation. *bioRxiv*, page 705095, 2019.

[60] Mengnan Du, Ninghao Liu, and Xia Hu. Techniques for Interpretable Machine Learning. *Communications of the ACM*, 63(1):68–77, 2019.

[61] Feng Duan, Dongxue Lin, Wenyu Li, and Zhao Zhang. Design of a Multimodal EEG-based Hybrid BCI System with Visual Servo Module. *ransactions on Autonomous Mental Development*, 7(4):332–341, 2015.

[62] Günter Edlinger and Christoph Guger. Can Dry EEG Sensors Improve the Usability of SMR, P300 and SSVEP Based BCIs? In *Towards Practical Brain-Computer Interfaces*, pages 281–300. 2012.

[63] Erdem Erkan and Mehmet Akbaba. A study on Performance Increasing in SSVEP based BCI Application. *Engineering Science and Technology, an International Journal*, 21(3):421–427, 2018.

[64] Christina Farmaki, Georgios Christodoulakis, and Vangelis Sakkalis. Applicability of SSVEP-based Brain-computer Interfaces for Robot Navigation in Real Environments. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2768–2771. IEEE, 2016.

[65] Muhamed Farooq and Omid Dehzangi. High Accuracy Wearable SSVEP Detection using Feature Profiling and Dimensionality Reduction. In *International Conference on Wearable and Implantable Body Sensor Networks*, pages 161–164. IEEE, 2017.

[66] Oliver Faust, Yuki Hagiwara, Tan Jen Hong, Oh Shu Lih, and U Rajendra Acharya. Deep Learning for Healthcare Applications based on Physiological Signals: A Review. *Computer methods and programs in biomedicine*, 161:1–13, 2018.

[67] Siamac Fazli, Cristian Grozea, Márton Danóczy, Benjamin Blankertz, Florin Popescu, and Klaus-Robert Müller. Subject Independent EEG-based BCI Decoding. In *Advances in Neural Information Processing Systems*, pages 513–521, 2009.

[68] SM Fernandez-Fraga, MA Aceves-Fernandez, JC Pedraza-Ortega, and S Tovar-Arriaga. Feature Extraction of EEG Signal upon BCI Systems Based on Steady-State Visual Evoked Potentials Using the Ant Colony Optimization Algorithm. *Discrete Dynamics in Nature and Society*, 2018, 2018.

[69] Álvaro Fernández-Rodríguez, Francisco Velasco-Álvarez, and Ricardo Ron-Angevin. Review of Real Brain-controlled Wheelchairs. *Journal of neural engineering*, 13(6):061001, 2016.

[70] Andrea Finke, Nils Hachmeister, Hannes Riechmann, and Helge Ritter. Thought-controlled Robots - Systems, Studies and Future Challenges. *International Conference on Robotics and Automation*, pages 3403–3408, 2013.

[71] Andrea Finke, Benjamin Rudgalwis, Holger Jakusch, and Helge Ritter. Towards Multi-user Brain-robot Interfaces for Humanoid Robot Control. *IEEE-RAS International Conference on Humanoid Robots*, pages 532–537, 2012.

[72] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. GAN-based Synthetic Medical Image Augmentation for Increased CNN Performance in Liver Lesion Classification. *Neurocomputing*, 321:321–331, 2018.

[73] Leo Galway, Paul McCullagh, Gaye Lightbody, Chris Brennan, and David Trainor. The Potential of the Brain-computer Interface for Learning: A Technology Review. In *International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, pages 1554–1559. IEEE, 2015.

[74] Qiang Gao, Lixiang Dou, Abdelkader Nasreddine Belkacem, and Chao Chen. Noninvasive Electroencephalogram Based Control of a Robotic Arm for Writing Task using Hybrid BCI System. *BioMed Research International*, 2017, 2017.

[75] Gaetano Gargiulo, Rafael A Calvo, Paolo Bifulco, Mario Cesarelli, Craig Jin, Armin Mohamed, and André van Schaik. A New EEG Recording System for Passive Dry Electrodes. *Clinical Neurophysiology*, 121(5):686–693, 2010.

[76] Felix Gembler, Piotr Stawicki, Abdul Saboor, and Ivan Volosyak. Dynamic Time Window Mechanism for Time Synchronous VEP-based BCIs—Performance Evaluation with a Dictionary-supported BCI Speller Employing SSVEP and c-VEP. *PloS one*, 14(6):e0218177, 2019.

[77] Emmanouil Giannakakis, Frances Hutchings, Christoforos A Papasavvas, Cheol E Han, Bernd Weber, Chencheng Zhang, and Marcus Kaiser. Computational Modelling of the Long-term Effects of Brain Stimulation on the Local and Global Structural Connectivity of Epileptic Patients. *Plos one*, 15(2):e0221380, 2020.

[78] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):142–158, 2016.

[79] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, Physiotoolkit, and Physionet. *Circulation*, 101(23):e215–e220, 2000.

[80] Stepan Gomilko, Alina Zimina, and Evgeny Shandarov. Attention Training Game with Aldebaran Robotics NAO and Brain-Computer Interface. In *International Conference on Interactive Collaborative Robotics*, volume 545, pages 27–31. Springer, 2016.

[81] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[82] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[83] David Gouaillier, Vincent Hugel, Pierre Blazevic, Chris Kilner, Jerome Monceaux, Pascal Lafourcade, Brice Marnier, Julien Serre, and Bruno Maisonnier. The nao humanoid: a combination of performance and affordability. *CoRR abs/0807.3223*, 2008.

[84] David Gouaillier, Vincent Hugel, Pierre Blazevic, Chris Kilner, Jérôme Monceaux, Pascal Lafourcade, Brice Marnier, Julien Serre, and Bruno Maisonnier. Mechatronic Design of NAO Humanoid. In *International Conference Robotics and Automation*, pages 769–774. IEEE, 2009.

[85] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved Training of Wasserstein GANs. In *Advances in Neural Information Processing Systems*, pages 5769–5779, 2017.

[86] Arzu Güneysu and H Levent Akin. An SSVEP based BCI to Control a Humanoid Robot by using Portable EEG Device. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6905–6908. IEEE, 2013.

[87] Ting Han, Lichao Xiu, and Guoming Yu. The Impact of Media Situation on People's Memory Effect – an ERP Study. *Computers in Human Behavior*, page 106180, 2019.

[88] Xu Han, Ke Lin, Shangkai Gao, and Xiaorong Gao. A Novel System of SSVEP-based Human–robot Coordination. *Journal of neural engineering*, 16(1):016006, 2018.

[89] Kay Gregor Hartmann, Robin Tibor Schirrmeister, and Tonio Ball. EEG-GAN: Generative Adversarial Networks for Electroencephalograhic (EEG) Brain Signals. *arXiv preprint arXiv:1806.01875*, 2018.

[90] Muhammad Haziq Hasbulah, Fairul Azni Jafar, and Mohd Hisham Nordin. Fundamental of Electroencephalogram (EEG) Review for Brain-Computer Interface (BCI) System. *International Research Journal of Engineering and Technology*, 2019.

[91] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Conference Computer Vision and Pattern Recognition*. IEEE, 2016.

[92] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786):504–507, 2006.

[93] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-term Memory. *Neural computation*, 9(8):1735–1780, 1997.

[94] Steven CH Hoi, Doyen Sahoo, Jing Lu, and Peilin Zhao. Online Learning: A Comprehensive Survey. *arXiv preprint arXiv:1802.02871*, 2018.

[95] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, 13(2):415–425, 2002.

[96] Hong Hu, Jing Zhao, Hongbo Li, Wei Li, and Genshe Chen. Telepresence control of Humanoid Robot via High-frequency Phase-tagged SSVEP Stimuli. *IEEE International Workshop on Advanced Motion Control*, (61473207):214–219, 2016.

[97] Ji-Young Hwang, Min-Ho Lee, and Seong-Whan Lee. A Brain-computer Interface Speller using Peripheral Stimulus-based SSVEP and P300. In *International Winter Conference on Brain-Computer Interface (BCI)*, pages 77–78. IEEE, 2017.

[98] Sunhee Hwang, Kibeom Hong, Guiyoung Son, and Hyeran Byun. EZSL-GAN: EEG-based Zero-Shot Learning Approach using a Generative Adversarial Network. In *International Winter Conference Brain-Computer Interface*, pages 1–4, 2019.

[99] Tommi Jaakkola and David Haussler. Exploiting Generative Models in Discriminative Classifiers. In *Advances in Neural Information Processing Systems*, pages 487–493, 1999.

[100] Jun Jiang, An Wang, Yu Ge, and Zongtan Zhou. Brain-actuated Humanoid Robot Control using One Class Motor Imagery Task. *Chinese Automation Congress*, pages 587–590, 2013.

[101] Sunil Kalagi, José Machado, Vitor Carvalho, Filomena Soares, and Demétrio Matos. Brain Computer Interface Systems using Non-invasive Electroencephalogram Signal: A literature Review. In *International Conference on Engineering, Technology and Innovation*, pages 1578–1583. IEEE, 2017.

[102] Emmanuel K Kalunga, Sylvain Chevallier, Quentin Barthélemy, Karim Djouani, Eric Monacelli, and Yskandar Hamam. Online SSVEP-based BCI using Riemannian Geometry. *Neurocomputing*, 191:55–68, 2016.

[103] Bojan Kerous, Filip Skola, and Fotis Liarokapis. EEG-based BCI and Video Games: A Progress Report. *Virtual Reality*, 22(2):119–135, 2018.

[104] Byungju Kim, Hyunwoo Kim, Kyungsu Kim, Sungjin Kim, and Junmo Kim. Learning Not to Learn: Training Deep Neural Networks with Biased Data. In *IEEE Conference Computer Vision and Pattern Recognition*, pages 9012–9020, 2019.

[105] Jeehoon Kim, Jeongsu Lee, Chungmin Han, and Kwangsuk Park. An Instant Donning Multi-Channel EEG Headset (with Comb-Shaped Dry Electrodes) and BCI Applications. *Sensors*, 19(7):1537, 2019.

[106] Seong Gon Kim, Nawanol Theera-Ampornpunt, Chih-Hao Fang, Mrudul Harwani, Ananth Grama, and Somali Chaterji. Opening Up the Blackbox: An Interpretable Deep Neural Network-based Classifier for Cell-type Specific Enhancer Predictions. *BMC systems biology*, 10(2):243–258, 2016.

[107] Diederik P Kingma and Jimmy Ba. Adam: A method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[108] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[109] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.

[110] Siddharth Kohli and Alexander J. Casson. Towards Out-of-the-lab EEG in Uncontrolled Environments: Feasibility Study of Dry EEG Recordings During Exercise Bike Riding. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015-Novem:1025–1028, 2015.

[111] R. Kottaimalai, M. Pallikonda Rajasekaran, V. Selvam, and B. Kannapiran. EEG signal classification using Principal Component Analysis with Neural Network in Brain Computer Interface applications. *IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology*, (Iceccn):227–231, 2013.

[112] Matthias Krauledat, Michael Tangermann, Benjamin Blankertz, and Klaus-Robert Müller. Towards Zero Training for Brain-Computer Interfacing. *PloS one*, 3(8):e2967, 2008.

[113] Andrea Kübler. The History of BCI: From A Vision for the Future to Real Support for Personhood in People with Locked-in Syndrome. *Neuroethics*, pages 1–18, 2019.

[114] Mikolaj E Kundegorski and Toby P Breckon. A Photogrammetric Approach for Real-time 3D Localization and Tracking of Pedestrians in Monocular Infrared Imagery. In *Optics and Photonics for Counterterrorism, Crime Fighting, and Defence X; and Optical Materials and Biomaterials in Security and Defence Systems Technology XI*, volume 9253, page 92530I. International Society for Optics and Photonics, 2014.

[115] No-Sang Kwak, Klaus-Robert Müller, and Seong-Whan Lee. A Convolutional Neural Network for Steady State Visual Evoked Potential Classification Under Ambulatory Environment. *PloS one*, 12(2):e0172578, 2017.

[116] No-Sang Kwak, Dong-Ok Won, Keun-Tae Kim, Hee-Jin Park, and Seong-Whan Lee. Analysis of Steady State Visual Evoked Potentials based on Viewing Distance Changes for Brain-machine Interface Speller. In *International Conference on Systems, Man, and Cybernetics*, pages 001502–001505. IEEE, 2016.

[117] Zirui Lan, Olga Sourina, Lipo Wang, Reinhold Scherer, and Gernot R Müller-Putz. Domain Adaptation Techniques for EEG-based Emotion Recognition: A Comparative Study on Two Public Datasets. *IEEE Transactions on Cognitive and Developmental Systems*, 11(1):85–94, 2018.

[118] Ka Fai Lao, Chi Man Wong, Ze Wang, and Feng Wan. Learning Prototype Spatial Filters for Subject-Independent SSVEP-Based Brain-Computer Interface. In *International Conference on Systems, Man, and Cybernetics*, pages 485–490. IEEE, 2018.

[119] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, Hugo Larochelle, and Ole Winther. Autoencoding beyond Pixels using a Learned Similarity Metric. In *International Conference Machine Learning*, pages 1558–1566, 2016.

[120] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. EEGNet: A Compact Convolutional Neural Network for EEG-based Brain–computer Interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

[121] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

[122] PoLei Lee, HsiangChih Chang, TsungYu Hsieh, HuaTing Deng, and ChiaWei Sun. A Brain-wave-actuated Small Robot Car using Ensemble Empirical Mode Decomposition-based Approach. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 42(5):1053–1064, 2012.

[123] Robert Leeb, Luca Tonin, Martin Rohm, Lorenzo Desideri, Tom Carlson, and Jose del R Millan. Towards Independence: a BCI Telepresence Robot for People with Severe Motor Disabilities. *Proceedings of the IEEE*, 103(6):969–982, 2015.

[124] Damien Lesenfants, Dina Habbal, Z Lugo, M Lebeau, Petar Horki, E Amico, Christoph Pokorny, F Gomez, A Soddu, Gernot Müller-Putz, et al. An Independent SSVEP-based Brain–computer Interface in Locked-in Syndrome. *Journal of neural engineering*, 11(3):035002, 2014.

[125] Jinpeng Li, Shuang Qiu, Yuan-Yuan Shen, Cheng-Lin Liu, and Huiguang He. Multisource Transfer Learning for Cross-Subject EEG Emotion Recognition. *IEEE Transactions on Cybernetics*, 2019.

[126] Mengfan Li, Wei Li, Jing Zhao, Qinghao Meng, Fuchun Sun, and Genshe Chen. An Adaptive P300 Model for Controlling a Humanoid Robot with Mind. *IEEE International Conference on Robotics and Biomimetics*, (December):1390–1395, 2013.

[127] Muyang Li, Ji Lin, Yaoyao Ding, Zhijian Liu, Jun-Yan Zhu, and Song Han. Gan Compression: Efficient Architectures for Interactive Conditional GANs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5284–5294, 2020.

[128] Wei Li, Yunyi Li, Genshe Chen, Qinghao Meng, Ming Zeng, and Fuchun Sun. Acquiring Brain Signals of Imagining Humanoid Robot Walking Behavior via Cerebot. In *Foundations*

and Practical Applications of Cognitive Systems and Information Processing, pages 617–627. Springer, 2014.

[129] Lun De Liao, Chi Yu Chen, I Jan Wang, Sheng Fu Chen, Shih Yu Li, Bo Wei Chen, Jyh Yeong Chang, and Chin Teng Lin. Gaming Control using a Wearable and Wireless EEG-based Brain-computer Interface Device with Novel Dry Foam-based Sensors. *Journal of NeuroEngineering and Rehabilitation*, 9(January):5, 2012.

[130] Lun De Liao, Shang Lin Wu, Chang Hong Liou, Shao Wei Lu, Shi An Chen, Sheng Fu Chen, Li Wei Ko, and Chin Teng Lin. A Novel 16-Channel Wireless System for Electroencephalography Measurements with Dry Spring-loaded Sensors. *IEEE Transactions on Instrumentation and Measurement*, 63(6):1545–1555, 2014.

[131] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision*, pages 740–755. Springer, 2014.

[132] Yuan-Pin Lin, Yijun Wang, and Tzyy-Ping Jung. A Mobile SSVEP-based Brain-computer Interface for Freely Moving Humans: The Robustness of Canonical Correlation Analysis to Motion Artifacts. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1350–1353. IEEE, 2013.

[133] Yuan-Pin Lin, Yijun Wang, Chun-Shu Wei, and Tzyy-Ping Jung. Assessing the Quality of Steady-State Visual-Evoked Potentials for Moving Humans using a Mobile Electroencephalogram Headset. *Frontiers in Human Neuroscience*, 8(March):1–10, 2014.

[134] Zachary C Lipton, John Berkowitz, and Charles Elkan. A Critical Review of Recurrent Neural Networks for Sequence Learning. *arXiv preprint arXiv:1506.00019*, 2015.

[135] Giuseppe Lisi, Masashi Hamaya, Tomoyuki Noda, and Jun Morimoto. Dry-wireless EEG and Asynchronous Adaptive Feature Extraction Towards a Plug-and-play Co-adaptive Brain Robot Interface. In *International Conference on Robotics and Automation*, pages 959–966. IEEE, 2016.

[136] Quan Liu, Kun Chen, Qingsong Ai, and Sheng Quan Xie. Recent Development of Signal Processing Algorithms for SSVEP-based Brain Computer Interfaces. *Journal Medical and Biological Engineering*, 34(4):299–309, 2014.

[137] Shaolin Liu, Fei Wang, Shichao Wu, Yahui Zhang, Ying Wei, Wei Wu, Haobo Zhao, and Yuze Zhang. Research of Mobile Robot Control System Based on SSVEP Brain Computer Interaction. In *Chinese Control And Decision Conference*. IEEE, 2018.

[138] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. SSD: Single Shot Multibox Detector. In *European Conference on Computer Vision*, pages 21–37. Springer, 2016.

[139] M. A. Lopez-Gordo, D. Sanchez-Morillo, and F. Pelayo Valle. Dry EEG Electrodes. *Sensors*, 14(7):12847–12870, 2014.

[140] Fabien Lotte. Generating Artificial EEG Signals to Reduce BCI Calibration Time. In *International Brain-Computer Interface Workshop*, pages 176–179, 2011.

[141] Fabien Lotte. Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-based Brain–Computer Interfaces. *Proceedings of the IEEE*, 103(6):871–890, 2015.

[142] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. A Review of Classification Algorithms for EEG-based Brain–Computer Interfaces: a 10 Year Update. *Journal Neural Engineering*, 15(3):031005, 2018.

[143] Fabien Lotte, Laurent Bougrain, and Maureen Clerc. Electroencephalography (EEG)-based Brain-Computer Interfaces. In *Wiley Encyclopedia of Electrical and Electronics Engineering*, page 44. Wiley, 2015.

[144] Fabien Lotte, Anatole Lécuyer, and Cuntai Guan. Towards a Fully Interpretable EEG-based BCI System. 2010.

[145] Yun Luo. EEG Data Augmentation for Emotion Recognition Using a Conditional Wasserstein GAN. In *International Conference Engineering in Medicine and Biology Society*, pages 2535–2538, 2018.

[146] Robert Gabriel Lupu, Florina Ungureanu, and Corina Cimpanu. Brain-Computer Interface: Challenges and Research Perspectives. In *International Conference on Control Systems and Computer Science*, pages 387–394. IEEE, 2019.

[147] Jiaxin Ma, Yu Zhang, Yunjun Nam, Andrzej Cichocki, and Fumitoshi Matsuno. EOG/ERP Hybrid Human-machine Interface for Robot Control. *IEEE International Conference on Intelligent Robots and Systems*, pages 859–864, 2013.

[148] Tyler C. Major and James M. Conrad. A Survey of Brain Computer Interfaces and Their Applications. *Conference Proceedings - IEEE SOUTHEASTCON*, 2014.

[149] Christian Mandel, Thorsten Lüth, Tim Laue, Thomas Röfer, Axel Gräser, and Bernd Krieg-Brückner. Navigating a smart wheelchair with a brain-computer interface interpreting steady-state visual evoked potentials. In *International Conference on Intelligent Robots and Systems*, pages 1118–1125. IEEE, 2009.

[150] Nikolay V Manyakov, Nikolay Chumerin, Arne Robben, Adrien Combaz, Marijn van Vliet, and Marc M Van Hulle. Sampled Sinusoidal Stimulation Profile and Multichannel Fuzzy Logic Classification for Monitor-based Phase-coded SSVEP Brain–computer Interfacing. *Journal of neural engineering*, 10(3):036011, 2013.

[151] Xiaoqian Mao, Mengfan Li, Wei Li, Linwei Niu, Bin Xian, Ming Zeng, and Genshe Chen. Progress in EEG-based Brain Robot Interaction Systems. *Computational Intelligence and Neuroscience*, 2017, 2017.

[152] Francesco Marini, Clement Lee, Johanna Wagner, Scott Makeig, and Mateusz Gola. A Comparative Evaluation of Signal Quality Between a Research-grade and a Wireless Dry-electrode Mobile EEG System. *Journal of neural engineering*, 16(5):054001, 2019.

[153] David Marshall, Damien Coyle, Shane Wilson, and Michael Callaghan. Games, Gameplay, and BCI: The State of the Art. *IEEE Transactions on Computational Intelligence and AI in Games*, 5(2):82–99, 2013.

[154] Kyle E Mathewson, Tyler JL Harrison, and Sayeed AD Kizuk. High and Dry? Comparing Active Dry EEG Electrodes to Active and Passive Wet Electrodes. *Psychophysiology*, 54(1):74–82, 2017.

[155] Dennis J. McFarland, William A. Sarnacki, George Townsend, Theresa Vaughan, and Jonathan R. Wolpaw. The P300-based Brain-computer Interface (BCI): Effects of Stimulus Rate. *Clinical Neurophysiology*, 122(4):731–737, 2011.

[156] DJ McFarland and JR Wolpaw. EEG-based Brain–computer Interfaces. *current opinion in Biomedical Engineering*, 4:194–200, 2017.

[157] Ramtin Mehraram, Marcus Kaiser, Ruth Cromarty, Sara Graziadio, John T O'Brien, Alison Killen, John-Paul Taylor, and Luis R Peraza. Weighted Network Measures Reveal Differences Between Dementia Types: An EEG Study. *Human Brain Mapping*, 41(6):1573–1590, 2020.

[158] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*, 2013.

[159] Jdel R Millan. On the Need for On-Line Learning in Brain-Computer Interfaces. In *International Joint Conference Neural Networks*, pages 2877–2882, 2004.

[160] Jesus Minguillon, M Angel Lopez-Gordo, and Francisco Pelayo. Trends in EEG-BCI for Daily-life: Requirements for Artifact Removal. *Biomedical Signal Processing and Control*, 31:407–418, 2017.

[161] Swapnil Mishra, Siraj M Tamboli, and Rohit Rajaraam Patil. A Review on Futuristic Technology "Brain Computer Interface (BCI)". *Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org*, 7(4), 2017.

[162] Tim Mullen, Christian Kothe, Yu Mike Chi, Alejandro Ojeda, Trevor Kerth, Scott Makeig, Gert Cauwenberghs, and Tzyy-Ping Jung. Real-time Modeling and 3D Visualization of Source Dynamics and Connectivity using Wearable EEG. In *International Conference of Engineering in Medicine and Biology Society*, pages 2184–2187. IEEE, 2013.

[163] Tim R Mullen, Christian AE Kothe, Yu Mike Chi, Alejandro Ojeda, Trevor Kerth, Scott Makeig, Tzyy-Ping Jung, and Gert Cauwenberghs. Real-time Neuroimaging and Cognitive Monitoring using Wearable Dry EEG. *IEEE Transactions on Biomedical Engineering*, 62(11):2553–2567, 2015.

[164] Sandra Mara Torres Müller, Teodiano Freire Bastos, and Mário Sarcinelli Filho. Proposal of a SSVEP-BCI to Command a Robotic Wheelchair. *Journal of Control, Automation and Electrical Systems*, 24(1-2):97–105, 2013.

[165] Gernot R Müller-Putz, René Riedl, and Selina C Wriessnegger. Electroencephalography (EEG) as a Research Tool in the Information Systems Discipline: Foundations, Measurement, and Applications. *CAIS*, 37:46, 2015.

[166] Masaki Nakanishi, Yijun Wang, Yu-Te Wang, and Tzyy-Ping Jung. A Comparison Study of Canonical Correlation Analysis Based Methods for Detecting Steady-state Visual Evoked Potentials. *PloS one*, 10(10):e0140703, 2015.

[167] Masaki Nakanishi, Yijun Wang, Yu-Te Wang, Yasue Mitsukura, and Tzyy-Ping Jung. Generating Visual Flickers for Eliciting Robust Steady-state Visual Evoked Potentials at Flexible Frequencies using Monitor Refresh Rate. *PloS one*, 9(6):e99235, 2014.

[168] Masaki Nakanishi, Yu-Te Wang, Chun-Shu Wei, Kuan-Jung Chiang, and Tzyy-Ping Jung. Facilitating Calibration in High-Speed BCI Spellers via Leveraging Cross-Device Shared Latent Responses. *IEEE Transactions on Biomedical Engineering*, 2019.

[169] Samreen Naz and N Bawane. Recent Trends in BCI based Speller System: A Survey Report. *International Journal of Engineering Science*, 6(7), 2016.

[170] Hooman Nezamfar, Seyed Sadegh Mohseni Salehi, Matt Higger, and Deniz Erdogmus. Code-vep vs. eye tracking: A comparison study. *Brain sciences*, 8(7):130, 2018.

[171] W. C. Ng, H. L. Seet, K. S. Lee, N. Ning, W. X. Tai, M. Sutedja, J. Y H Fuh, and X. P. Li. Micro-spike EEG electrode and the Vacuum-casting Technology for Mass Production. *Journal of Materials Processing Technology*, 209(9):4434–4438, 2009.

[172] Trung-Hau Nguyen and Wan-Young Chung. A Single-Channel SSVEP-Based BCI Speller Using Deep Learning. *IEEE Access*, 7:1752–1763, 2018.

[173] Spiros Nikolopoulos. MAMEM EEG SSVEP Dataset I (256 channels, 11 subjects, 5 frequencies), Jan 2016.

[174] Anthony M Norcia, L Gregory Appelbaum, Justin M Ales, Benoit R Cottereau, and Bruno Rossion. The Steady-state Visual Evoked Potential in Vision Research: A Review. *Journal of vision*, 15(6):4–4, 2015.

[175] Ewan Nurse, Benjamin S Mashford, Antonio Jimeno Yepes, Isabell Kiral-Kornek, Stefan Harrer, and Dean R Freestone. Decoding EEG and LFP Signals using Deep Learning: heading TrueNorth. In *Proceedings of the ACM International Conference on Computing Frontiers*, pages 259–266. ACM, 2016.

[176] Augustus Odena, Christopher Olah, and Jonathon Shlens. Conditional Image Synthesis with Auxiliary Classifier GANs. In *International Conference Machine Learning*, pages 2642–2651, 2017.

[177] Vangelis P Oikonomou, Georgios Liaros, Kostantinos Georgiadis, Elisavet Chatzilari, Katerina Adam, Spiros Nikolopoulos, and Ioannis Kompatsiaris. Comparative Evaluation of State-of-the-art Algorithms for SSVEP-based BCIs. *arXiv preprint arXiv:1602.00904*, 2016.

[178] Ozan Özdenizci, Ye Wang, Toshiaki Koike-Akino, and Deniz Erdoğmuş. Adversarial deep learning in eeg biometrics. *IEEE Signal Processing Letters*, 26(5):710–714, 2019.

[179] Ozan Özdenizci, Ye Wang, Toshiaki Koike-Akino, and Deniz Erdoğmuş. Learning Invariant Representations From EEG via Adversarial Inference. *IEEE Access*, 8, 2020.

[180] Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, and Mubarak Shah. Generative Adversarial Networks Conditioned by Brain Signals. In *International Conference Computer Vision*, pages 3410–3418, 2017.

[181] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Zachary DeVito, Zeming Lin, Alban Desmaison, and Adam Lerer. Automatic Differentiation in PyTorch. In *Advances in Neural Information Processing Systems*, pages 1–4, 2017.

[182] Guangying Pei, Jinglong Wu, Duanduan Chen, Guoxin Guo, Shuozhen Liu, Mingxuan Hong, and Tianyi Yan. Effects of an Integrated Neurofeedback System with Dry Electrodes: EEG Acquisition and Cognition Assessment. *Sensors*, 18(10):3396, 2018.

[183] Jonathan Peirce, Jeremy R Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. PsychoPy2: Experiments in Behavior Made Easy. *Behavior research methods*, 51(1):195–203, 2019.

[184] Jonathan W Peirce. PsychoPy-Psychophysics Software in Python. *Journal of Neuroscience Methods*, 162(1-2):8–13, 2007.

[185] Chamika Janith Perera, Isira Naotunna, Chameera Sadaruwan, Ranathunga Arachchilage Ruwan Chandra Gopura, and Thilina Dulantha Lalitharatne. SSVEP based BMI for a Meal Assistance Robot. In *International Conference on Systems, Man, and Cybernetics*, pages 002295–002300. IEEE, 2016.

[186] Andreas Pinegger, Selina C Wriessnegger, Josef Faller, and Gernot R Müller-Putz. Evaluation of Different EEG Acquisition Systems Concerning Their Suitability for Building a Brain–computer Interface: Case Studies. *Frontiers in neuroscience*, 10:441, 2016.

[187] Joshua J Podmore, Toby P Breckon, Nik KN Aznan, and Jason D Connolly. On the Relative Contribution of Deep Convolutional Neural Networks for SSVEP-Based Bio-Signal Decoding in BCI Speller Applications. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(4):611–618, 2019.

[188] Narendra Prataksita, Yi Tseng Lin, Hung Chyun Chou, and Chung Hsien Kuo. Brain-Robot Control Interface: Development and Application. *IEEE International Symposium on Bioelectronics and Bioinformatics*, pages 10–13, 2014.

[189] Robert Prueckl and Christoph Guger. *A Brain-Computer Interface Based on Steady State Visual Evoked Potentials for Controlling a Robot*, pages 690–697. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[190] Aina Puce and Matti S Hämäläinen. A Review of Issues Related to Data Acquisition and Analysis in EEG/MEG Studies. *Brain sciences*, 7(6):58, 2017.

[191] Jie-Lin Qiu and Wei-Ye Zhao. Data Encoding Visualization Based Cognitive Emotion Recognition with AC-GAN Applied for Denoising. In *International Conference Cognitive Informatics & Cognitive Computing*, pages 222–227, 2018.

[192] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv preprint arXiv:1511.06434*, 2015.

[193] Rabie A Ramadan and Athanasios V Vasilakos. Brain Computer Interface: Control Signals Review. *Neurocomputing*, 223:26–44, 2017.

[194] Rajesh P.N. Rao. *Brain-Computer Interfacing: An Introduction*. Cambridge University Press, New York, NY, USA, 2013.

[195] Pedro Reis, Felix Hebenstreit, Florian Gabsteiger, Vinzenz von Tscharner, and Matthias Lochmann. Methodological Aspects of EEG and Body Dynamics Measurements During Motion. *Frontiers in human neuroscience*, 8:156, 2014.

[196] Angela I Renton, Jason B Mattingley, and David R Painter. Optimising Non-invasive Brain-computer Interface Systems for Free Communication Between naïve Human Participants. *Scientific reports*, 2019.

[197] Aya Rezeika, Mihaly Benda, Piotr Stawicki, Felix Gembler, Abdul Saboor, and Ivan Volosyak. Brain–computer Interface Spellers: A Review. *Brain sciences*, 8(4):57, 2018.

[198] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. In *International Conference Machine Learning*, pages 1278–1286, 2014.

[199] Hannes Riechmann, Andrea Finke, and Helge Ritter. Using a cVEP-Based Brain-Computer Interface to Control a Virtual Agent. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(6):692–699, 2016.

[200] Sujit Roy, Karl McCreadie, and Girijesh Prasad. Can a Single Model Deep Learning Approach Enhance Classification Accuracy of an EEG-based Brain-Computer Interface? In *IEEE International Conference on Systems, Man and Cybernetics*, 2019.

[201] Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep Learning-based Electroencephalography Analysis: A Systematic Review. *Journal of neural engineering*, 2019.

[202] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet Large Scale Visual Recognition Challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[203] Abdul Saboor, Felix Gembler, Mihaly Benda, Piotr Stawicki, Aya Rezeika, Roland Grichnik, and Ivan Volosyak. A Browser-driven SSVEP-based BCI Web Speller. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 625–630. IEEE, 2018.

[204] Batyrkhan Saduanov, Tohid Alizadeh, Jinung An, and Berdakh Abibullaev. Trained by Demonstration Humanoid Robot Controlled via a BCI System for Telepresence. In *International Conference on Brain-Computer Interface*, pages 1–4. IEEE, 2018.

[205] Ruslan Salakhutdinov. Learning Deep Generative Models. *Annual Review of Statistics and Its Application*, 2:361–385, 2015.

[206] Eduardo Santamaría-Vázquez, Víctor Martínez-Cagigal, Javier Gomez-Pilar, and Roberto Hornero. Asynchronous Control of ERP-Based BCI Spellers Using Steady-State Visual Evoked Potentials Elicited by Peripheral Stimuli. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(9):1883–1892, 2019.

[207] Prasanna Sattigeri, Samuel C Hoffman, Vijil Chenthamarakshan, and Kush R Varshney. Fairness GAN. *arXiv preprint arXiv:1805.09910*, 2018.

[208] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep Learning with Convolutional Neural Networks for EEG Decoding and Visualization. *Human Brain Mapping*, 38(11):5391–5420, 2017.

[209] Alois Schlögl. Outcome of the BCI-Competition 2003 on the Graz Data Set. *Berlin, Germany: Graz University of Technology*, 2003.

[210] Shili Sheng, Peipei Song, Lingyue Xie, Zhendong Luo, Wennan Chang, Shurui Jiang, Haoyong Yu, Chi Zhu, Jeffrey Too Chuan Tan, and Feng Duan. Design of an SSVEP-based BCI System With Visual Servo Module for a Service Robot to Execute Multiple Tasks. In *International Conference on Robotics and Automation*, pages 2267–2272. IEEE, 2017.

[211] Xiaokang Shu, Lin Yao, Jianjun Meng, Xinjun Sheng, and Xiangyang Zhu. Visual Stimulus Background Effects on SSVEP-Based BCI Towards a Practical Robot Car Control. *International Journal of Humanoid Robotics*, 12(02):1550014, 2015.

[212] Rajesh Singla and BA Haseena. Comparison of SSVEP Signal Classification Techniques using SVM and ANN Models for BCI Applications. *International Journal of Information and Electronics Engineering*, 4(1):6, 2014.

[213] Chris Solomon and Toby Breckon. *Fundamentals of Digital Image Processing*. Wiley, 2011.

[214] Rossella Spataro, Antonio Chella, Brendan Allison, Marcello Giardina, Rosario Sorbello, Salvatore Tramonte, Christoph Guger, and Vincenzo La Bella. Reaching and Grasping a Glass of Water by Locked-in ALS Patients Through a BCI-controlled Humanoid Robot. *Frontiers in human neuroscience*, 11:68, 2017.

[215] Ghislain St-Yves and Thomas Naselaris. Generative Adversarial Networks Conditioned on Brain Activity Reconstruct Seen Images. In *International Conference on Systems, Man, and Cybernetics*, pages 1054–1061. IEEE, 2018.

[216] Philipp Stankevich and Vladimir Spitsyn. A review of Brain-Computer Interface Technology. In *International Siberian Conference on Control and Communications*, pages 1–6. IEEE, 2015.

[217] Piotr Stawicki, Felix Gembler, Cheuk Yin Chan, Mihaly Benda, Aya Rezeika, Abdul Saboor, Roland Grichnik, and Ivan Volosyak. SSVEP-based BCI in Virtual Reality-control of a Vacuum Cleaner Robot. In *International Conference on Systems, Man, and Cybernetics*, pages 534–537. IEEE, 2018.

[218] Piotr Stawicki, Felix Gembler, and Ivan Volosyak. Driving a Semiautonomous Mobile Robotic Car Controlled by an SSVEP-Based BCI. *Computational Intelligence and Neuroscience*, 2016, 2016.

[219] Piotr Stawicki, Felix Gembler, and Ivan Volosyak. Driving a sSemiautonomous Mobile Robotic Car Controlled by an SSVEP-based BCI. *Computational intelligence and neuroscience*, 2016:5, 2016.

[220] Sebastian Stober, Avital Sternin, Adrian M Owen, and Jessica A Grahn. Deep Feature Learning for EEG Recordings. *arXiv preprint arXiv:1511.04306*, 2015.

[221] Irene Sturm, Sebastian Lapuschkin, Wojciech Samek, and Klaus-Robert Müller. Interpretable Deep Neural Networks for Single-trial EEG Classification. *Journal of neuroscience methods*, 274:141–145, 2016.

[222] Mingui Sun, Wenyan Jia, Wei Liang, and Robert J. Sclabassi. A Low-impedance, Skin-grabbing, and Gel-free EEG Electrode. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 1992–1995, 2012.

[223] Yousef Rezaei Tabar and Ugur Halici. A Novel Deep Learning Approach for Classification of EEG Motor Imagery Signals. *Journal of Neural Engineering*, 14(1):016003, 2017.

[224] Jingshen Tang, Jun Jiang, Yang Yu, and Zongtan Zhou. Humanoid Robot Operation by a Brain-computer Interface. *Proceedings - 2015 7th International Conference on Information Technology in Medicine and Education, ITME 2015*, pages 476–479, 2016.

[225] Richard MG Tello, Sandra MT Müller, Teodiano Bastos-Filho, and Andre Ferreira. A Comparison of Techniques and Technologies for SSVEP Classification. In *ISSNIP-IEEE Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living*, pages 1–6. IEEE, 2014.

[226] John Thomas, Tomasz Maszczyk, Nishant Sinha, Tilmann Kluge, and Justin Dauwels. Deep Learning-Based Classification for Brain-Computer Interfaces. In *International Conference on Systems, Man, and Cybernetics*, pages 234–239. IEEE, 2017.

[227] Emmanuele Tidoni, Pierre Gergondet, Gabriele Fusco, Abderrahmane Kheddar, and Salvatore M Aglioti. The Role of Audio-Visual Feedback in a Thought-based Control of a Humanoid Robot: A BCI Study in Healthy and Spinal Cord Injured People. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(6):772–781, 2017.

[228] Devashree Tripathy and Jagdish Lal Raheja. Design and Implementation of Brain Computer Interface Based Robot Motion Control. *Advances in Intelligent Systems and Computing*, 328:289–296, 2015.

[229] Goran Udovičić, Ante Topić, and Mladen Russo. Wearable Technologies for Smart Environments: A Review with Emphasis on BCI. In *2016 24th International Conference on Software, Telecommunications and Computer Networks*, pages 1–9. IEEE, 2016.

[230] Swati Vaid, Preeti Singh, and Chamandeep Kaur. EEG Signal Analysis for BCI Interface: A Review. *International Conference on Advanced Computing and Communication Technologies*, 2015-April:143–147, 2015.

[231] Albert Vilamala, Kristoffer H Madsen, and Lars K Hansen. Deep Convolutional Neural Networks for Interpretable Analysis of EEG Sleep Stage Scoring. In *International Workshop on Machine Learning for Signal Processing*, pages 1–6. IEEE, 2017.

[232] Hanh Vu, Bonkon Koo, and Seungjin Choi. Frequency Detection for SSVEP-based BCI Using Deep Canonical Correlation Analysis. In *International Conference on Systems, Man, and Cybernetics*, pages 001983–001987. IEEE, 2016.

[233] Li Wang, Xiaochu Liu, Zhongwei Liang, Zhao Yang, and Xiao Hu. Analysis and Classification of Hybrid BCI Based on Motor Imagery and Speech Imagery. *Measurement*, 147:106842, 2019.

[234] Yijun Wang, Xiaogang Chen, Xiaorong Gao, and Shangkai Gao. A Benchmark Dataset for SSVEP-based Brain–computer Interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1746–1752, 2016.

[235] Nicholas Waytowich, Vernon J Lawhern, Javier O Garcia, Jennifer Cummings, Josef Faller, Paul Sajda, and Jean M Vettel. Compact Convolutional Neural Networks for Classification of Asynchronous Steady-state Visual Evoked Potentials. *Journal of neural engineering*, 15(6):066031, 2018.

[236] Ethan Webster, Hadi Habibzadeh, James JS Norton, Theresa M Vaughan, and Tolga Soyata. An Unsupervised Channel-Selection Method for SSVEP-based BCI Systems. In *Annual Ubiquitous Computing, Electronics & Mobile Communication Conference*, pages 626–632. IEEE, 2018.

[237] Karl Weiss, Taghi M. Khoshgoftaar, and DingDing Wang. *A Survey of Transfer Learning*, volume 3. Springer International Publishing, 2016.

[238] Benjamin Wittevrongel and Marc M Van Hulle. Hierarchical Online SSVEP Spelling Achieved with Spatiotemporal Beamforming. In *Statistical Signal Processing Workshop*, pages 1–5. IEEE, 2016.

[239] SC Wriessnegger, A Pinegger, and GR Mueller-Putz. The Evaluation of Different EEG Sensor Technologies. In *Information Systems and Neuroscience*, pages 85–90. Springer, 2015.

[240] Chung-Min Wu, Yeou-Jiunn Chen, Ilham AE Zaeni, and Shih-Chung Chen. A New SSVEP based BCI Application on the Mobile Robot in a Maze Game. In *International Conference on Advanced Materials for Science and Engineering*, pages 550–553. IEEE, 2016.

[241] Eric Wu, Kevin Wu, David Cox, and William Lotter. Conditional Infilling GANs for Data Augmentation in Mammogram Classification. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pages 98–106. Springer, 2018.

[242] Xiaoting Wu, Li Zheng, Lu Jiang, Xiaoshan Huang, Yuanyuan Liu, Lihua Xing, Xiao Xing, Yijun Wang, Weihua Pei, Xiaowei Yang, et al. A Dry Electrode Cap and Its Application in a Steady-State Visual Evoked Potential-Based Brain–Computer Interface. *Electronics*, 8(10):1080, 2019.

[243] Xiao Xing, Yijun Wang, Weihua Pei, Xuhong Guo, Zhiduo Liu, Fei Wang, Gege Ming, Hongze Zhao, Qiang Gui, and Hongda Chen. A High-speed SSVEP-based BCI using Dry EEG Electrodes. *Scientific reports*, 8(1):14708, 2018.

[244] Depeng Xu, Shuhan Yuan, Lu Zhang, and Xintao Wu. FairGAN: Fairness-Aware Generative Adversarial Networks. In *International Conference Big Data*, pages 570–575, 2018.

[245] Florian Yger, Fabien Lotte, and Masashi Sugiyama. Averaging Covariance Matrices for EEG Signal Classification based on the CSP: An Empirical Study. In *European Signal Processing Conference*, pages 2721–2725. IEEE, 2015.

[246] Erwei Yin, Zongtan Zhou, Jun Jiang, Yang Yu, and Dewen Hu. A Dynamically Optimized SSVEP Brain-computer Interface (BCI) Speller. *IEEE Transactions on Biomedical Engineering*, 62(6):1447–1456, 2014.

[247] Ahmed Youssef Ali Amer, Benjamin Wittevrongel, and Marc Van Hulle. Accurate Decoding of Short, Phase-Encoded SSVEPs. *Sensors*, 18(3):794, 2018.

[248] Ting Yu, Chun-Shu Wei, Kuan-Jung Chiang, Masaki Nakanishi, and Tzyy-Ping Jung. EEG-Based User Authentication Using a Convolutional Neural Network. In *International Conference Neural Engineering*, pages 1011–1014. IEEE, 2019.

[249] Thorsten O Zander, Lena M Andreessen, Angela Berg, Maurice Bleuel, Juliane Pawlitzki, Lars Zawallich, Laurens R Krol, and Klaus Gramann. Evaluation of a Dry EEG System for Application of Passive Brain-computer Interfaces in Autonomous Driving. *Frontiers in human neuroscience*, 11:78, 2017.

[250] Rosanne Zerafa, Tracey Camilleri, Owen Falzon, and Kenneth P Camilleri. To Train or Not To Train? A Survey on Training of Feature Extraction Methods for SSVEP-based BCIs. *Journal of neural engineering*, 15(5):051001, 2018.

[251] Liming Zhang, Xiaodong Zhang, Zhufeng Lu, and Rui Li. A New Object-oriented SSVEP-based BCI Paradigm Using Continuous Action Scene. In *International Conference on CYBER Technology in Automation, Control, and Intelligent Systems*, pages 1078–1082. IEEE, 2017.

[252] Nannan Zhang, Jun Jiang, Jingsheng Tang, Zongtan Zhou, and Dewen Hu. A Novel Steady-State Visually Evoked Potential-Based Brain-Computer-Interface Paradigm to Steer a Humanoid Robot. In *International Conference on Instrumentation and Measurement, Computer, Communication and Control*, pages 774–778. IEEE, 2014.

[253] Qiqi Zhang and Ying Liu. mproving Brain Computer Interface Performance by Data Augmentation with Conditional Deep Convolutional Generative Adversarial Networks. *arXiv preprint arXiv:1806.07108*, 2018.

[254] Yu Zhang, Guoxu Zhou, Jing Jin, Xingyu Wang, and Andrzej Cichocki. Frequency Recognition in SSVEP-based BCI using Multiset Canonical Correlation Analysis. *International journal of neural systems*, 24(04):1450013, 2014.

[255] Yubin Zhang, Jun Xie, Guanghua Xu, Peng Fang, Guangjing Du, Xiaodong Zhang, Min Li, Sicong Zhang, Tangfei Tao, and Hua Yuan. Performance Evaluation of a "Switch-To-Target" Based Asynchronous SSVEP BCI Paradigm. In *International Conference on Ubiquitous Robots*, pages 32–37. IEEE, 2019.

[256] Jing Zhao, Wei Li, and Mengfan Li. Comparative Study of SSVEP- and P300-based Models for the Telepresence Control of Humanoid Robots. *PLoS ONE*, 10(11):1–18, 2015.

[257] Jing Zhao, Wei Li, Xiaoqian Mao, Hong Hu, Linwei Niu, and Genshe Chen. Behavior-based SSVEP Hierarchical Architecture for Telepresence Control of Humanoid Robot to Achieve Full Body Movement. *IEEE Transactions on Cognitive and Developmental Systems*, 8920(2013):1–1, 2016.

[258] Jing Zhao, Wei Li, Xiaoqian Mao, Hong Hu, Linwei Niu, and Genshe Chen. Behavior-based SSVEP Hierarchical Architecture for Telepresence Control of Humanoid Robot to Achieve Full-body Movement. *IEEE Transactions on Cognitive and Developmental Systems*, 9(2):197–209, 2017.

[259] Jing Zhao, Wei Li, Xiaoqian Mao, and Mengfan Li. SSVEP-based Experimental Procedure for Brain-Robot Interaction with Humanoid Robots. *Journal of Visualized Experiments*, (105), 2015.

[260] Jing Zhao, Qinghao Meng, Wei Li, Mengfan Li, and Genshe Chen. SSVEP-based Hierarchical Architecture for Control of a Humanoid Robot with Mind. *Proceedings of the World Congress on Intelligent Control and Automation*, 2015-March(March):2401–2406, 2015.

[261] Jing Zhao, Qinghao Meng, Wei Li, Mengfan Li, Fuchun Sun, and Genshe Chen. An OpenViBE-based Brainwave Control System for Cerebot. *IEEE International Conference on Robotics and Biomimetics*, (December):1169–1174, 2013.

[262] Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, 7(3):162–175, 2015.

Appendix A

Data Acquisition

Cognionics data acquisition software is the software that enables to stream and record the bio-signals in real-time from the dry-EEG device by using Bluetooth technology. The software is a Windows-based software as in Figure A.1. The software gives the flexibility to alter headset settings according to the experimenting nature, such as the sampling rate and the selection of sensors used. The information on the connected device is displayed on the primary window of the software. The signals can be saved using the software by recording the data, or they can be streamed using different data streaming protocols such as; Lab Streaming Layer (LSL) or streaming using Remote Data Access (RDA), to be analysed in different software.

Figure A.2 shows sensors available on Quick-20 headset, where non-required channels can be disabled. This option is useful for experiments which do not require all the sensors, or for a quick calibration and setup phase. This software is also useful for checking the impedance value and the connectivity for each sensor channel to assure the quality of the signals throughout the experimental protocol (Figure A.3).

RDA is one of an integrated streaming protocols provided by the Cognionics data acquisition software to share the signals over the network in real-time that provides the software remote access to transfer data, either locally or in another machine over the network. Algorithm 2 details the communication process between the Cognionics data acquisition software run on a

Figure A.1: Main interface of the Cognionics Data Acquisition software

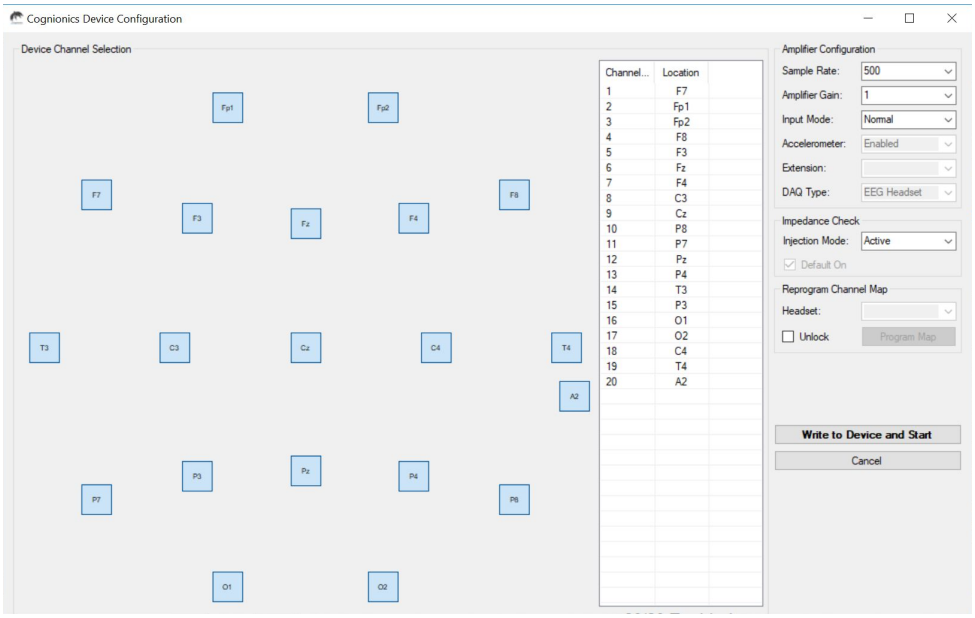


Figure A.2: Pre-experimentation configuration including sensor selection

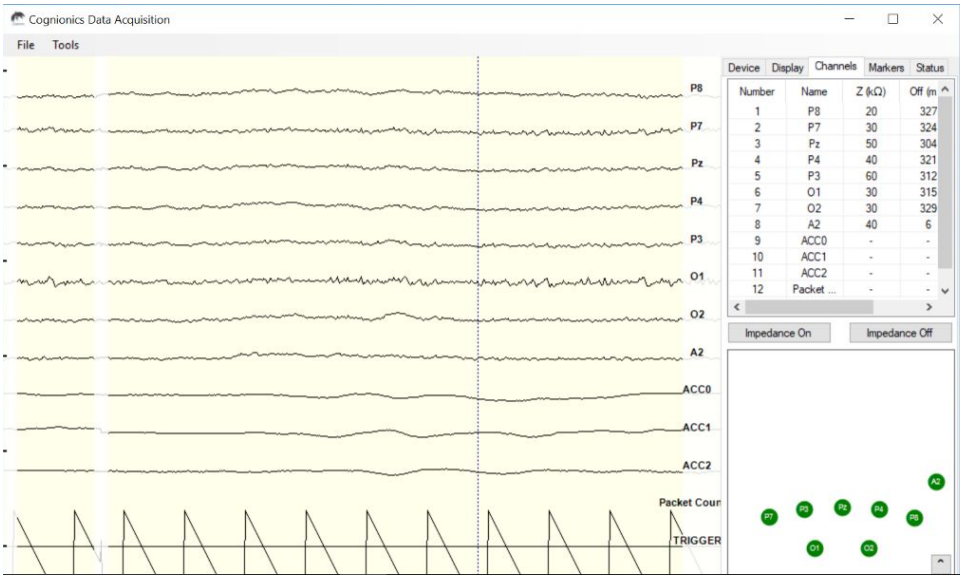


Figure A.3: Real-time signals streaming with real-time impedance value for each sensor

Algorithm 2: Streaming bio-signals from Cognionics data acquisition

Input : Frequency.
Output: Signals array.

```
Initialise Cognionics machine as client          // IP address and RDA port = 51244
Create buffer window and size for client
Connect to client
Start streaming
while c < n_t do
    run windowflip
    pull the required sample from RDA buffer
end while
Add sample to array
Save the array and the ground truth (frequency)
```

host computer (Microsoft Windows), with the main Python platform running SSVEP stimuli and RDA client protocol on a host computer (Linux).

In order to stream and record the bio-signals of the subjects performing SSVEP tasks, we setup integration between all the software and hardware components as in Figure 3.3. The SSVEP

stimuli is displayed in the client machine (primary machine in this work) where a Python program is used as the main program on the machine to display the stimuli and stream the bio-signals from the server machine via RDA. While having fixation upon the stimuli, the bio-signals are streamed via dry-EEG headset connected to the data acquisition software running on the client machine. The software is responsible for sending over the signal packets through the network via the RDA client package to the primary machine.

Appendix B

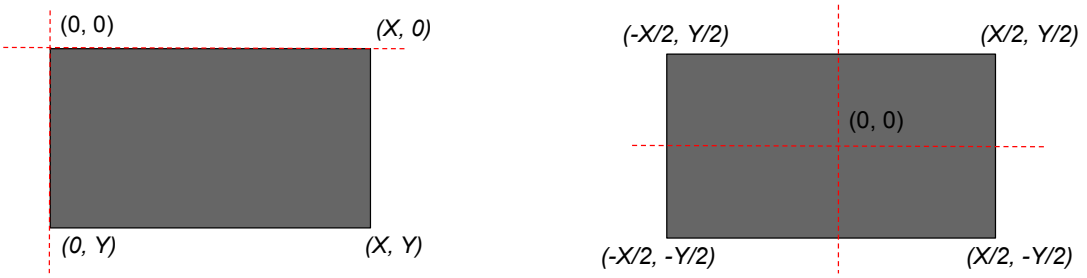
Variable Stimuli

Our variable stimuli employ SSD network to predict the objects in the scene by using OpenCV to display the video frame, and to draw and display the bounding box [28]. In order to display the predicted objects as SSVEP stimuli, some modifications are required in order to display the prediction of SSD on the Psychopy window. Algorithm 3 details the process of getting a frame at a time, then flickering the detected object according to the predetermined frequency chosen for that object. As mentioned, if the SSD network detects an object being present in the current frame, the bounding box points and the predicted class label are returned.

We manipulate this returned information from the SSD predictor to draw the stimuli via Psychopy, We calculate the new box array for the bounding box points to revise the bounding box array returned in OpenCV format. The new array according to Psychopy is drawn in such format as in Figure B.1b as compared to OpenCV format in Figure B.1a.

Algorithm 3: Displaying variable stimuli with flickering bounding boxes on psychopy via SSD

Input : Video frame
Output: Variable Stimuli
Call SSD predictor function
Return class index and bounding box array points ; `// in OpenCV format`
if *array points is bigger than 0* **then**
    Calculate the new bounding box array points ; `// in Psychopy format`
    Set the position and size of each square of the detected object according to new points
    Check index class to appoint frame array accordingly
    Use the same way in Algorithm 1 to display the stimuli
else
    Do nothing
end if



(a) OpenCV

(b) Psychopy

Figure B.1: The coordinate format for window display for both software, where (0, 0) is the origin and (X, Y) is the width and height of the monitor resolution used.